ASPECTS OF MULTIVARIATE CONTROL CHARTS

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

by

Francis Bernard Alt

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in the School of Industrial and Systems Engineering

Georgia Institute of Technology

December, 1973

ASPECTS OF MULTIVARIATE CONTROL CHARTS

Approved:

D. C. Montgomery, Chairman

H. M. Wadsworth

J. J. Goode

Date approved by Chairman: 19 Nov. 1973

ACKNOWLEDGMENTS

Dr. D. C. Montgomery deserves my gratitude for serving as my thesis advisor. He has provided numerous comments.

Dr. H. M. Wadsworth patiently waded through several formative versions of the thesis as a member of my reading committee. His conscientious and helpful attitude will long be appreciated and remembered.

I wish to extend special thanks to Dr. J. J. Goode who made contributions "above and beyond the call of duty" both as a member of my reading committee and as an educator. His availability and excellent mathematical ability have always been appreciated.

Although many people deserve my thanks for their direct contributions, several people must be explicitly mentioned. Dr. M. S. Bazaraa has my sincere thanks for his friendship and assistance with parts of Chapters IV and VI. Dr. J. J. Jarvis also was very helpful with parts of Chapter IV. Dr. H. Chernoff of Stanford deserves credit for his assistance with resolutions of the paradox mentioned in Chapter VI.

There are also several people who must be specifically mentioned for their indirect contributions. Dr. A. J. Duncan of The Johns Hopkins University first introduced me to the field of statistics and encouraged me to work in this area. Dr. J. W. Walker further fostered my interest in this area and deserves special gratitude for having always been a source of inspiration. Dr. J. A. Harrison of Morehouse spent considerable time explaining to me the rudiments of simultaneous inference and

and other aspects of multivariate statistics. Dr. W. W. Hines deserves special thanks for his continuing encouragement and for providing every possible assistance. Dr. R. N. Lehrer has been most dependable in always providing me with some form of financial assistance. Dean S. C. Webb and Dr. S. W. Thornton have both been great sources of encouragement.

Many fellow graduate students deserve my gratitude for their friendship, inspiration and assistance. Special mention must be made of Mr. Robert L. Bulfin, Jr., who spent considerable time with the development of the necessary computer programs. Mr. Jerry T. Hagerman also deserves special mention for his excellent artwork.

Special thanks also goes to Mrs. Betty Sims for her enduring and friendly attitude throughout the somewhat magical task of transforming a handwritten copy into an excellent final typewritten draft.

Finally, I wish to thank my wife, Judith, and my daughter, Elizabeth, for patiently sacrificing time and attention which they deserved. They have always provided their love, regardless.

This thesis is dedicated to my parents who dedicated their lives to the raising of their three children.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

In the sequel, a $(p\times 1)$ vector of random variables $X_1,X_2,\ldots,X_p$ will be denoted by an upper case letter underscored by a tilde: $\underset{\sim}{X}$. A $(p\times 1)$ vector of constants $a_1,a_2,\ldots,a_p$ will be denoted by a lower case letter underscored by a tilde: $\underset{\sim}{a}$. A matrix will be denoted by an upper case letter as will a univariate random variable; the distinction between the two should be explicit from the context.

The general multivariate statistical quality control problem considers a repetitive process where each item is characterized by p quality characteristics, $X_1,X_2,\ldots,X_p$, which are random variables because of the chance causes inherent in the process. It will be assumed that the probability law, denoted by $f_{\underset{\sim}{X}}(\underset{\sim}{x})$, associated with the p quality characteristics is multivariate normal with population mean $\underset{\sim}{\mu}$ and covariance matrix Σ. Thus, in a repetitive manufacturing operation, multiple measurements will be made on certain of the successively manufactured items, as opposed to the single measurement per item recorded in the univariate quality control problem. In any repetitive process, it is desired that the multiple measurements made on each of the items behave as though they were obtained from a population having specified or nominal values of $\underset{\sim}{\mu}$ and Σ, denoted by $\underset{\sim}{\mu}_0$ and Σ_0, respectively. When changes in the process cause elements of $\underset{\sim}{\mu}$ or Σ to shift from the nominal values, it becomes necessary to detect these changes to insure a uniform quality product.

If there is only one quality characteristic X, where X is normally distributed with mean μ_0 and variance σ_0^2 $\{$denoted $X \sim N(\mu_0, \sigma_0^2)\}$, the multivariate problem reduces to the well-known univariate problem. Refer to Duncan (18). When there is a $(p \times 1)$ vector $\underset{\sim}{X}$ of quality characteristics, where $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}_0, \Sigma_0)$, Hotelling (36,37), Jackson (40,41) and others have treated certain aspects of this problem.

For p = 1, the control charts are usually given by a central line and upper and lower control limits. When the population mean and variance are specified, a random sample of size n is selected and sample statistics are computed for controlling the process mean and variability. For successive random samples, this can be viewed as repeated tests of significance. That is, the decision maker is successively testing

$$H_0: \mu = \mu_0 \quad \text{vs.} \quad H_1: \mu \neq \mu_0$$

and

$$H_0: \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_1: \sigma^2 \neq \sigma_0^2.$$

Because of this relationship to tests of significance, the upper and lower control limits could be replaced by percentage points which reflect the regions of rejection of the H_0. This hypothesis testing viewpoint permits the tractability necessary for $p > 1$. For p = 1 with the population mean and variance unspecified, the usual ad hoc procedure is to determine the unbiased estimates of the population parameters and to use these estimates in place of the population parameters in the control limits. A similar procedure will be used for $p > 1$.

Hotelling adopted the hypothesis testing viewpoint in controlling the process mean for $p > 1$ with the population parameters specified by briefly mentioning the use of a χ^2 statistic. His generalization (35) of Student's ratio enables one to test the multivariate hypotheses

$$H_0: \underset{\sim}{\mu} = \underset{\sim}{\mu}_0 \quad \text{vs.} \quad H_1: \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0,$$

when Σ is not known. In his 1947 paper, he briefly mentions how this generalization, denoted T^2, could be used in the quality control problem of testing bombsights. He also suggested that a generalized statistic, $T_0^2 = \sum_h T_h^2$, be plotted against time, as illustrated in Figure 1.

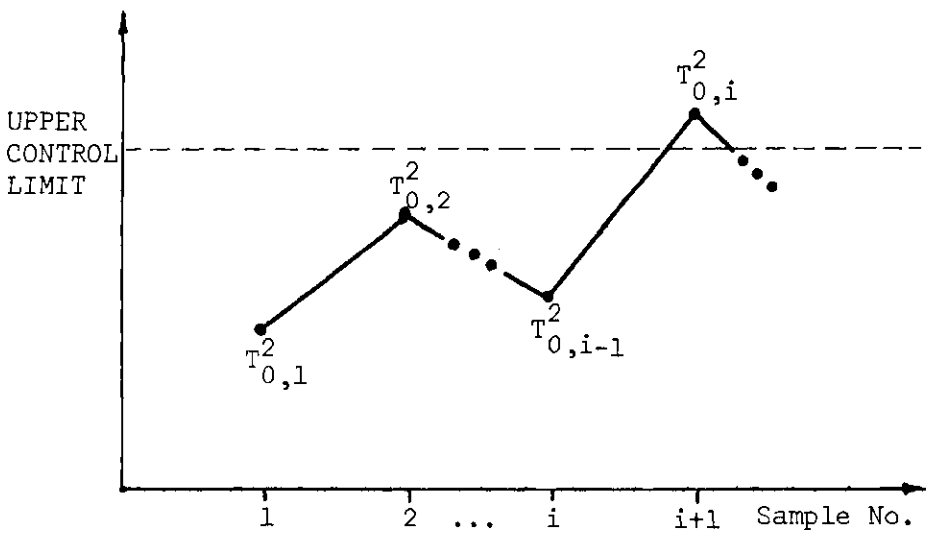

Figure 1. Hotelling's T^2 Chart

In his 1951 paper, he generalized his previous work and obtained the distribution of T_0^2 for $p = 2$. Jackson also concentrated on shifts in the mean vector and his 1956 paper deals exclusively with the use of the T^2 statistic for p = 2. He also prefers to treat the control of individual observations rather than sample means as is usually done for p = 1. Jackson suggests that the control region be displayed as an ellipse as illustrated in Figure 2. His 1959 paper generalizes his previous work and introduces the concept of principal components as a measure of dispersion control. Both Hotelling's and Jackson's works will be elaborated upon in later chapters after the necessary background has been introduced.

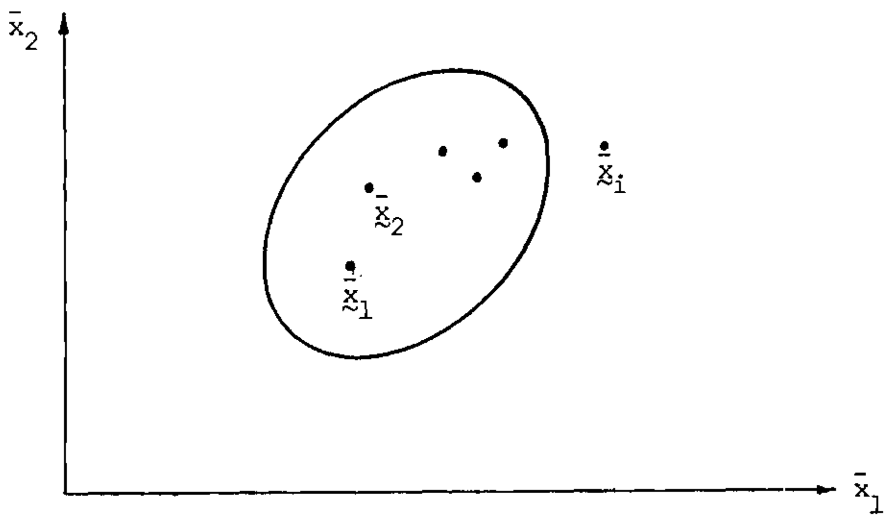

Figure 2. Elliptical Control Chart for p = 2

For more than one quality characteristic, the all too frequent and incorrect approach is to attempt to control the p-related variables by p univariate procedures. Unfortunately, this drastically alters the probability of incorrectly rejecting the null hypothesis.

The general objectives and direction of this research are as follows:

1. Present in a precise format the statistical methodology used in multivariate statistical quality control. Although some knowledge of univariate statistics and matrix algebra is assumed, the material is intended to be relatively self-contained and this results in the inclusion of Chapters II and III. This objective has resulted in a direct-product proof of the independence of the sample mean vector and the sample covariance matrix.

2. Develop and elaborate on control charts for the mean of the process when there is more than one quality characteristic. Chapter IV considers the case when the population parameters are unknown, while Chapter V treats the cases when the population mean or covariance matrix are unknown. In conjunction with this, simultaneous techniques are presented for the determination of those quality characteristics which are out of control. Guidelines will be presented for determining which technique to use.

3. Quite frequently, the χ^2 statistic is used in place of the T^2 statistic. In view of this, certain aspects of both of these statistics will be presented. As a result, two resolutions will be offered concerning a paradox that arose in looking at the noncentrality parameters of the T^2 and χ^2 distributions.

4. In the case of one quality characteristic, the range chart or sigma chart is most often used to control process variability. Multivariate analogues will be presented, namely, the sample generalized variance and the multivariate range.

CHAPTER II

THE MULTIVARIATE NORMAL AND RELATED PROPERTIES

In dealing with the simultaneous control of p related variables, the data consist of a set of (p×1) vectors of measurements. For the statistical analysis of such data, the mathematical model assumed is that each data point is an observation from a multivariate normal probability distribution. Some justification for using this model is now in order.

1. Justification for Use of the Multivariate Normal

One of the most compelling reasons for using this model is the large number of situations to which it is amenable. Francis Galton, the geneticist, made use of this model in the 1880's when he found that human height tends to "regress" back to type. In a multitude of problems since then, the multivariate normal distribution has been found to be a sufficiently close approximation to the underlying probability distribution of the population.

Another justification for using the multivariate normal distribution stems from the multidimensional version of the Central-Limit Theorem. The multidimensional version asserts that, under certain conditions, the sum of a large number of p-dimensional random vectors is asymptotically distributed as a p-variate normal. For the exact conditions regarding the validity of the theorem, see Cramér (15). Thus, since multiple measurements are often sums of small independent effects, the multivariate central-limit theorem further justifies the use of the multivariate normal probability distribution.

In the univariate case, the assumption of an underlying normal population facilitates the derivation of certain test statistics. This facility carries over to the multidimensional problem and further justifies the use of the multivariate normal distribution as the probability distribution of the population from which samples will be taken.

2. The Density of the Multivariate Normal

The general multivariate normal distribution will be seen as very similar to the general univariate normal distribution. In the univariate case, a random variable is said to be normally distributed if it can be expressed as an affine transformation of a standard normal random variable Z, where $f_Z(z) = (2\pi)^{-1/2}\exp\left(-(1/2)z^2\right)$, for $-\infty < z < \infty$. Thus, any variable of the form $X = rZ + \mu$, where $r \neq 0$, is normally distributed. If E and V denote the expected value and variance operators, respectively, then $E(X) = \mu$ and $V(X) = r^2$. Usually, r is taken as positive since -Z is standard normal if Z is standard normal. The development of the multivariate normal model as suggested by Tucker (68) and others will follow the presentation of three necessary lemmas.

Lemma 2.1. If the transformation from the vector $\underset{\sim}{z}$ to the vector $\underset{\sim}{x}$ is given by $\underset{\sim}{z} = B\underset{\sim}{x}$ where B is a nonsingular $(p\times p)$ matrix, then $\mathrm{mod}(|B|) = \mathrm{mod}\ J$, where $\mathrm{mod}(|B|)$ denotes the absolute value of the determinant of B and mod J denotes the absolute value of the Jacobian of the transformation.

Proof. By definition, the Jacobian is the determinant of the matrix $B = [b_{ij}]$ where $\partial z_i/\partial x_j = b_{ij}$. That is, $|[\partial z_i/\partial x_j]| = |[b_{ij}]| = |B|$. Hence, $\mathrm{mod}(|[\partial z_i/\partial x_j]|) = \mathrm{mod}(|B|)$. ||

Lemma 2.2. If B is a ($p \times p$) nonsingular matrix, then $(B')^{-1} = (B^{-1})'$. That is, the inverse of the transpose of B equals the transpose of the inverse of B.

Proof. Since $BB^{-1} = I$, $(BB^{-1})' = I'$, and $(B^{-1})'B' = I$. Multiplying on the right by $(B')^{-1}$ yields the desired result. ||

Lemma 2.3. If B is a nonsingular ($p \times p$) matrix, and if $\underset{\sim}{x}$ is a ($p \times 1$) vector, then $\underset{\sim}{x} = \underset{\sim}{0}$ if and only if $B\underset{\sim}{x} = \underset{\sim}{0}$.

Proof. If $\underset{\sim}{x} = \underset{\sim}{0}$, then $B\underset{\sim}{x} = \underset{\sim}{0}$. If $B\underset{\sim}{x} = \underset{\sim}{0}$, then $B^{-1}B\underset{\sim}{x} = \underset{\sim}{0}$ and $\underset{\sim}{x} = \underset{\sim}{0}$. ||

Definition 2.1. The ($p \times 1$) vector $\underset{\sim}{X}$ is said to have a multivariate normal distribution if there exist p independent standard normal random variables $Z_1, \ldots, Z_p$, a ($p \times 1$) vector of constants $\underset{\sim}{\mu} = [\mu_1, \ldots, \mu_p]'$, and a nonsingular ($p \times p$) matrix $R = [r_{ij}]$ such that $\underset{\sim}{X} = R\underset{\sim}{Z} + \underset{\sim}{\mu}$.

The joint probability density of $\underset{\sim}{X}$ can now be derived based on the above definition. Since the Z_h are independent random variables and $f_{Z_h}(z_h) = (2\pi)^{-1/2}\exp\left(-(1/2)z_h^2\right)$, then

$$f_{\underset{\sim}{Z}'}(\underset{\sim}{z}') = (2\pi)^{-p/2}\exp\left(-(1/2)\underset{\sim}{z}'\underset{\sim}{z}\right).$$

Consider the transformation $\underset{\sim}{z} = R^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})$. From Lemma 2.1, mod J = mod($|R^{-1}|$), and

$$f_{\underset{\sim}{X}'}(\underset{\sim}{x}') = (2\pi)^{-p/2}\exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'(R^{-1})'R^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]\mathrm{mod}(|R^{-1}|).$$

Define the matrix $\Sigma = RR'$. Then $\Sigma^{-1} = (RR')^{-1} = (R')^{-1}R^{-1} = (R^{-1})'R^{-1}$, by the use of Lemma 2.2. If B and C are (p×p) matrices, then $|BC| = |B||C|$ and $|B'| = |B|$. Hence, $|\Sigma^{-1}| = |(R^{-1})'R^{-1}| = |R^{-1}||R^{-1}| = |R^{-1}|^2$, and $\mathrm{mod}(|R^{-1}|) = |\Sigma^{-1}|^{1/2} = |\Sigma|^{-1/2}$. Thus, the probability density of $\underset{\sim}{X}$ is:

$$f_{\underset{\sim}{X}'}(\underset{\sim}{x}') = (2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-(1/2)(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\}. \qquad (1)$$

However, it remains to determine the significance of $\underset{\sim}{\mu}$ and Σ^{-1}.

3. The Expected Value and Covariance Matrix and Related Properties

If B is a (p×p) matrix, the expected value of B, denoted $\mathcal{E}B$, is the expected value of each element of B. That is, $\mathcal{E}B = [\mathcal{E}b_{ij}]$. The (p×p) variance-covariance matrix of a (p×1) vector $\underset{\sim}{X}$, denoted by $\mathcal{C}(\underset{\sim}{X},\underset{\sim}{X}')$, is $\mathcal{E}(\underset{\sim}{X}-\underset{\sim}{\mu})(\underset{\sim}{X}-\underset{\sim}{\mu})'$. It is assumed that $\mathcal{E}\underset{\sim}{X}$ and $\mathcal{C}(\underset{\sim}{X},\underset{\sim}{X}')$ always exist.

Theorem 2.1. If the (p×1) vector $\underset{\sim}{X}$ has a multivariate normal distribution with density

$$f_{\underset{\sim}{X}'}(\underset{\sim}{x}') = (2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-(1/2)(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\},$$

then $\mu = E\underset{\sim}{X}$ and $\Sigma = C(\underset{\sim}{X},\underset{\sim}{X}')$.

Proof. Since $\underset{\sim}{X}$ has a multivariate normal distribution, there exists a nonsingular matrix R such that $\underset{\sim}{X} = R\underset{\sim}{Z} + \underset{\sim}{\mu}$. Then, $E\underset{\sim}{X} = RE\underset{\sim}{Z} + \underset{\sim}{\mu}$. But $E\underset{\sim}{Z} = [EZ_h] = [0] = \underset{\sim}{0}$ since $Z_h \sim N(0,1)$ for $h = 1,\ldots,p$. Thus, $E\underset{\sim}{X} = R\underset{\sim}{0} + \underset{\sim}{\mu}$, or $E\underset{\sim}{X} = \underset{\sim}{\mu}$.

Let $d\underset{\sim}{x}'$ denote $dx_1 dx_2 \ldots dx_p$. Then,

$$C(\underset{\sim}{X},\underset{\sim}{X}') = E(\underset{\sim}{X}-\underset{\sim}{\mu})(\underset{\sim}{X}-\underset{\sim}{\mu})' = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\underset{\sim}{x}-\underset{\sim}{\mu})(\underset{\sim}{x}-\underset{\sim}{\mu})' f_{\underset{\sim}{X}'}(\underset{\sim}{x}')d\underset{\sim}{x}'.$$

Consider the transformation $(\underset{\sim}{x}-\underset{\sim}{\mu}) = R\underset{\sim}{z}$. Then, mod J = mod$(|R|) = |\Sigma|^{1/2}$, since $|\Sigma| = |R|^2$. In the exponent, $R'\Sigma^{-1}R = I$ since $\Sigma^{-1} = (R')^{-1}R^{-1}$. Thus,

$$C(\underset{\sim}{X},\underset{\sim}{X}') = (2\pi)^{-p/2}|\Sigma|^{-1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} R\underset{\sim}{z}\underset{\sim}{z}'R'\exp\left(-(1/2)\underset{\sim}{z}'\underset{\sim}{z}\right)|\Sigma|^{1/2}d\underset{\sim}{z}'$$

$$= R\left\{(2\pi)^{-p/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \underset{\sim}{z}\underset{\sim}{z}'\exp\left(-(1/2)\underset{\sim}{z}'\underset{\sim}{z}\right)d\underset{\sim}{z}'\right\}R'$$

$$= R(E\underset{\sim}{Z}\underset{\sim}{Z}')R'.$$

Now $E\underset{\sim}{Z}\underset{\sim}{Z}'$ is a $(p\times p)$ matrix with entries EZ_iZ_j. For $i \neq j$, $EZ_iZ_j = EZ_iEZ_j = 0$, since Z_i and Z_j are independent for all i and j and $EZ_i = 0$. For $i = j$, $EZ_iZ_j = EZ_i^2 = VZ_i + (EZ_i)^2 = 1 + 0 = 1$. Hence, $E\underset{\sim}{Z}\underset{\sim}{Z}' = I$ and $C(\underset{\sim}{X},\underset{\sim}{X}') = RR' = \Sigma$. ‖

Thus the vector $\underset{\sim}{\mu}$ is the expected value of $\underset{\sim}{X}$, and the matrix Σ

is the covariance matrix of $\underset{\sim}{X}$. Some of the properties of Σ and Σ^{-1} will now be stated.

Theorem 2.2. Both Σ and Σ^{-1} are symmetric and positive definite.

Proof. Since $\Sigma' = (RR')' = RR' = \Sigma$, Σ is symmetric. Σ is positive definite if $\underset{\sim}{x}'\Sigma\underset{\sim}{x} > 0$ for all nonnull $\underset{\sim}{x}$. By definition, $\Sigma = RR'$, where R is nonsingular. Then for $\underset{\sim}{x} \neq \underset{\sim}{0}$, $\underset{\sim}{x}'\Sigma\underset{\sim}{x} = x'RR'\underset{\sim}{x} = (R'\underset{\sim}{x})'(R'\underset{\sim}{x}) > 0$, by the contrapositive of Lemma 2.3. Similarly, it can be shown that Σ^{-1} is symmetric and positive definite. ‖

Many authors, such as Anderson (3) and Graybill (25) state a different definition of the multivariate normal distribution. Their definition states that Σ^{-1} is only some positive definite matrix whose elements are constants, $\underset{\sim}{\mu}$ is a $(p\times 1)$ vector of constants, and $(2\pi)^{-p/2}|\Sigma|^{-1/2}$ is denoted by k where k is some positive constant. This definition is equivalent to Definition 2.1 since there exists a $(p\times p)$ nonsingular symmetric matrix $R^{1/2}$ such that $R^{1/2}R^{1/2} = R$. For a detailed explanation of this, the reader should consult Tucker (68).

To give some insight into Definition 2.1, consider the bivariate normal. For $\underset{\sim}{X} = [X_1,X_2]'$, $E\underset{\sim}{X} = [\mu_1,\mu_2]'$. The correlation coefficient between X_1 and X_2, denoted by ρ, is defined to be

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\,\sqrt{\sigma_{22}}}$$

where $\sigma_{12} = \sigma_{21}$ denotes the covariance between X_1 and X_2 and σ_{hh} denotes the variance of X_h for $h = 1,2$. Hence, $\sigma_{12} = \sigma_{21} = \rho\sqrt{\sigma_{11}}\,\sqrt{\sigma_{22}}$, and the covariance matrix for $\underset{\sim}{X}$ can be written as

$$\Sigma_{\underset{\sim}{X}} = \begin{bmatrix} E(X_1-\mu_1)^2 & E(X_1-\mu_1)(X_2-\mu_2) \\ E(X_2-\mu_2)(X_1-\mu_1) & E(X_2-\mu_2)^2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Obviously,

$$|\Sigma_{\underset{\sim}{X}}|^{-1/2} = (\sigma_1^2\sigma_2^2-\rho^2\sigma_1^2\sigma_2^2)^{-1/2} = \left(\sigma_1^2\sigma_2^2(1-\rho^2)\right)^{-1/2} = \sigma_1^{-1}\sigma_2^{-1}(1-\rho^2)^{-1/2},$$

and

$$\Sigma_{\underset{\sim}{X}}^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{bmatrix}.$$

Thus, the probability density function for $\underset{\sim}{X}'$ is

$$f_{\underset{\sim}{X}'}(\underset{\sim}{x}') = (2\pi\sigma_1\sigma_2)^{-1}(1-\rho^2)^{-1/2}\exp\left\{-(1/2)(1-\rho^2)^{-1}\left[(x_1-\mu_1)^2\sigma_1^{-2} - 2\rho(x_1-\mu_1)(x_2-\mu_2)\sigma_1^{-1}\sigma_2^{-1} + (x_2-\mu_2)^2\sigma_2^{-2}\right]\right\}.$$

In Definition 2.1, if $\underset{\sim}{Z} = [Z_1,Z_2]'$ with $Z_h \sim N(0,1)$, then one particular R matrix such that $\underset{\sim}{X} = R\underset{\sim}{Z} + \underset{\sim}{\mu}$ would be

$$R = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2(1-\rho^2)^{1/2} \end{bmatrix},$$

and $\underset{\sim}{X} = R\underset{\sim}{Z} + \underset{\sim}{\mu}$ becomes

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2(1-\rho^2)^{1/2} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}.$$

It is easily verified that $RR' = \Sigma_{\underset{\sim}{X}}$. The above-mentioned affine transformation enables the generation of realizations of a general bivariate normal random vector since computer routines exist for the generation of realizations from a univariate standard normal random variable. This technique will be used in the Appendix to generate data for the purpose of demonstrating the use of multivariate quality control charts.

4. Other Properties of the Multivariate Normal

The moment-generating function of the general univariate normal random variable X, denoted by $M_X(t)$, can be found to equal $\exp\left(\mu t + (1/2)\sigma^2 t^2\right)$ by using the definition $M_X(t) = E\left(\exp(tX)\right)$. Remark 2.1 will show that a similar result holds for the multivariate case.

Remark 2.1. Let the $(p\times 1)$ vector $\underset{\sim}{X}$ have a multivariate normal distribution as stated in Definition 2.1. Then $M_{\underset{\sim}{X}}(\underset{\sim}{t}) = \exp\left(\underset{\sim}{t}'\underset{\sim}{\mu} + (1/2)\underset{\sim}{t}'\Sigma\underset{\sim}{t}\right)$.

Proof. $$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = E\left(\exp(\underset{\sim}{t}'\underset{\sim}{X})\right) = E\left(\exp(\underset{\sim}{t}'\underset{\sim}{\mu})\exp(\underset{\sim}{t}'R\underset{\sim}{Z})\right)$$

$$= \exp(\underset{\sim}{t}'\underset{\sim}{\mu})E\left(\exp(\underset{\sim}{t}'R\underset{\sim}{Z})\right).$$

Let $\underset{\sim}{s}'$ denote the $(1\times p)$ vector $\underset{\sim}{t}'R$. Then,

$$\begin{aligned} M_{\underset{\sim}{X}}(\underset{\sim}{t}) &= \exp(\underset{\sim}{t}'\underset{\sim}{\mu})E\left(\exp(\underset{\sim}{s}'\underset{\sim}{Z})\right) \\ &= \exp(\underset{\sim}{t}'\underset{\sim}{\mu}) \prod_{h=1}^{p} \exp\left((1/2)s_h^2\right) \\ &= \exp(\underset{\sim}{t}'\underset{\sim}{\mu})\exp\left((1/2)\underset{\sim}{s}'\underset{\sim}{s}\right) \\ &= \exp\left(\underset{\sim}{t}'\underset{\sim}{\mu}+(1/2)\underset{\sim}{t}'\Sigma\underset{\sim}{t}\right), \text{ since } \Sigma = RR'. \quad \| \end{aligned}$$

The distribution of a matrix transformation of a p-variate normal random vector may now be determined.

Theorem 2.3. If $\underset{\sim}{Y}$ is a $(p\times 1)$ vector such that $\underset{\sim}{Y}\sim N(\underset{\sim}{\mu},\Sigma)$ and $\underset{\sim}{X} = B\underset{\sim}{Y}$ where B is a $(q\times p)$ matrix with $q \le p$, then the $(q\times 1)$ vector $\underset{\sim}{X}$ is normally distributed with $E(\underset{\sim}{X}) = B\underset{\sim}{\mu}$ and $C(\underset{\sim}{X},\underset{\sim}{X}') = B\Sigma B'$.

Proof. Let $\underset{\sim}{\mu}_{\underset{\sim}{X}}$ and $\Sigma_{\underset{\sim}{X}}$ denote the mean vector and covariance matrix of $\underset{\sim}{X}$, respectively. Then

$$\mu_{\underset{\sim}{X}} = EB\underset{\sim}{Y} = BE\underset{\sim}{Y} = B\underset{\sim}{\mu},$$

and

$$\Sigma_{\underset{\sim}{X}} = E(\underset{\sim}{X}-\underset{\sim}{\mu}_{\underset{\sim}{X}})(\underset{\sim}{X}-\underset{\sim}{\mu}_{\underset{\sim}{X}})' = E(B\underset{\sim}{Y}-B\underset{\sim}{\mu})(B\underset{\sim}{Y}-B\underset{\sim}{\mu})' = B\Sigma B'.$$

To show that $\underset{\sim}{X}$ is normally distributed, the moment-generating function of $\underset{\sim}{X}$ will be considered. From Remark 2.1, $M_{\underset{\sim}{Y}}(\underset{\sim}{t}) = \exp\{\underset{\sim}{t}'\underset{\sim}{\mu}+(1/2)\underset{\sim}{t}'\Sigma\underset{\sim}{t}\}$. Hence,

$$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = E\exp(\underset{\sim}{t}'B\underset{\sim}{Y}) = E\exp\{(B'\underset{\sim}{t})'\underset{\sim}{Y}\} = M_{\underset{\sim}{Y}}(B'\underset{\sim}{t}).$$

But,

$$M_{\underset{\sim}{Y}}(B'\underset{\sim}{t}) = \exp\{(B'\underset{\sim}{t})'\underset{\sim}{\mu}+(1/2)(B'\underset{\sim}{t})'\Sigma(B'\underset{\sim}{t})\} = \exp\{\underset{\sim}{t}'(B\underset{\sim}{\mu})+(1/2)\underset{\sim}{t}'(B\Sigma B')\underset{\sim}{t}\},$$

which is the moment-generating function of a q-variate normal with mean $B\underset{\sim}{\mu}$ and covariance matrix $B\Sigma B'$. ||

Another property that will be needed concerning the p-variate normal distribution is the conditional distribution of a $(q\times 1)$ vector $\underset{\sim}{X}_1$, $q < p$, given that the $(p-q\times 1)$ vector $\underset{\sim}{X}_2 = \underset{\sim}{x}_2$. Before the conditional distribution and its properties can be stated, two lemmas will be presented.

Lemma 2.4. Let the $(p\times p)$ matrix Σ and its inverse, Λ, be partitioned in conformity with the partitioning of Σ as follows:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix},$$

where

$$\Sigma_{11} = (\sigma_{ij}),\quad \Lambda_{11} = (\lambda_{ij}),\quad i=1,\ldots,q,\quad j=1,\ldots,q,$$

$$\Sigma_{12} = (\sigma_{ij}),\quad \Lambda_{12} = (\lambda_{ij}),\quad i=1,\ldots,q,\quad j=1,\ldots,p-q,$$

$$\Sigma_{21} = (\sigma_{ij}),\quad \Lambda_{21} = (\lambda_{ij}),\quad i=1,\ldots,p-q,\quad j=1,\ldots,q,$$

$$\Sigma_{22} = (\sigma_{ij}),\quad \Lambda_{22} = (\lambda_{ij}),\quad i=1,\ldots,p-q,\quad j=1,\ldots,p-q.$$

Then, provided all the inverses exist,

$$\Lambda_{11} = (\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1},\qquad \Lambda_{12} = -\Sigma_{11}^{-1}\Sigma_{12}\Lambda_{22} = -\Lambda_{11}\Sigma_{12}\Sigma_{22}^{-1}$$

$$\Lambda_{21} = -\Sigma_{22}^{-1}\Sigma_{21}\Lambda_{11} = -\Lambda_{22}\Sigma_{21}\Sigma_{11}^{-1}, \text{ and } \Lambda_{22} = (\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}$$

Proof. Since Λ is the inverse of Σ,

$$\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\begin{bmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{bmatrix} = \begin{bmatrix}I_q & 0\\ 0 & I_{p-q}\end{bmatrix},$$

and this leads to the following equations:

$$\Sigma_{11}\Lambda_{11} + \Sigma_{12}\Lambda_{21} = I_q,\quad \Sigma_{11}\Lambda_{12} + \Sigma_{12}\Lambda_{22} = 0,$$

$$\Sigma_{21}\Lambda_{11} + \Sigma_{22}\Lambda_{21} = 0, \quad \Sigma_{21}\Lambda_{12} + \Sigma_{22}\Lambda_{22} = I_{p-q},$$

from which the result immediately follows. ||

The relations expressed in Lemma 2.4 are symmetric in that Σ and Λ can be interchanged.

Lemma 2.5. For Σ partitioned as in Lemma 2.4 and provided that Σ_{11}^{-1} and Σ_{22}^{-1} exist, then

$$|\Sigma| = |\Sigma_{11}||\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| = |\Sigma_{22}||\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|.$$

Proof. Since Σ_{11} and Σ_{22} are nonsingular,

$$\begin{bmatrix} I_{p-q} & -\Sigma_{21}\Sigma_{11}^{-1} \\ 0 & I_q \end{bmatrix} \begin{bmatrix} \Sigma_{22} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{11} \end{bmatrix} \begin{bmatrix} I_{p-q} & 0 \\ -\Sigma_{11}^{-1}\Sigma_{12} & I_q \end{bmatrix} = \begin{bmatrix} \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} & 0 \\ 0 & \Sigma_{11} \end{bmatrix}$$

and taking the determinants of both sides reveals that $|\Sigma| = |\Sigma_{11}||\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|$. The other result is obtained by reversing the roles of Σ_{11} and Σ_{22}. ||

Theorem 2.4. Let the $(p\times 1)$ vector $\underset{\sim}{X}$, where $\underset{\sim}{X}\sim N(\underset{\sim}{\mu},\Sigma)$, be partitioned into the $(q\times 1)$ and $(p-q\times 1)$ vectors $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$, respectively. Also let

$$\begin{bmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \text{ and } \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

be the corresponding partitions of $\underset{\sim}{\mu}$, Σ, and $\Lambda = \Sigma^{-1}$, respectively. Then the conditional density of $\underset{\sim}{X}_1$, given that $\underset{\sim}{X}_2 = \underset{\sim}{x}_2$, is normally distributed with mean $\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)$ and covariance matrix $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Proof. By definition,

$$f_{\underset{\sim}{X}_1'|\underset{\sim}{X}_2'=\underset{\sim}{x}_2'}(\underset{\sim}{x}_1'|\underset{\sim}{x}_2') = \frac{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\left(-(1/2)(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right)}{(2\pi)^{-(p-q)/2}|\Sigma_{22}|^{-1/2}\exp\left(-(1/2)(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)\right)} .$$

Now $|\Sigma|^{-1/2}/|\Sigma_{22}|^{-1/2} = |\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|^{-1/2}$ by Lemma 2.5. The exponent in the numerator, when expressed in terms of the corresponding submatrices, becomes $[-(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)'\Lambda_{11}(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)-(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)'\Lambda_{12}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)-(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Lambda_{21}(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)-(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Lambda_{22}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)]/2$. When the term, $(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)$, is added to and subtracted from the bracketed expression, it becomes $[-\left((\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)+\Lambda_{11}^{-1}\Lambda_{12}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)\right)'\Lambda_{11}\left((\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)+\Lambda_{11}^{-1}\Lambda_{12}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)\right)-(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Lambda_{22}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)+(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)]/2$. By Lemma 2.4, and the fact that $\Lambda_{22} - \Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12} = \Sigma_{22}^{-1}$, it now becomes $[-\left((\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)-\Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)\right)'(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\left((\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)-\Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)\right)-(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)]/2$, where the last term in the bracketed expression cancels with the exponent in the denominator of the expression for the conditional density. The conclusion now follows. $\|$

Note that the matrix $\Sigma_{12}\Sigma_{22}^{-1}$ is frequently called the matrix of regression coefficients, and $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is frequently denoted by $\Sigma_{11.2}$.

To give some insight into the above theorem, consider the bivariate normal. Then, the conditional density of X_1 given $X_2 = x_2$ is $N(\mu_1+(\sigma_1/\sigma_2)\rho(x_2-\mu_2),\sigma_1^2(1-\rho^2))$. A similar result is obtained for X_2 given x_1.

The final property that will be needed is a special case of a result due to Crámer and Wold which asserts that the distribution of a p-dimensional random variable, $\underset{\sim}{X}$, is completely determined by the one-dimensional distributions $\underset{\sim}{a}'\underset{\sim}{X}$ for every nonnull fixed $\underset{\sim}{a}$.

Theorem 2.5. Let $\underset{\sim}{X}$ be a p-dimensional random vector and let $\underset{\sim}{a}$ be a $(p\times 1)$ vector of constants. Then $\underset{\sim}{X}$ has a p-variate normal distribution if and only if $\underset{\sim}{a}'\underset{\sim}{X}$ has a univariate normal distribution for all nonnull $\underset{\sim}{a}$.

Proof. The proof follows immediately from the properties of the moment-generating function. ||

It should again be emphasized that the properties of the normal presented in this chapter are not new. Their reiteration in this chapter serves only to provide a background for the sequel and as an attempt to make this publication as self-contained as possible.

CHAPTER III

ESTIMATORS OF $\underset{\sim}{\mu}$ AND Σ AND DISTRIBUTIONAL PROPERTIES

The preceding discussion of the multivariate normal distribution assumed that $\underset{\sim}{\mu}$ and Σ were known. Since this is seldom true, this chapter will consider the derivation of the maximum likelihood estimators of $\underset{\sim}{\mu}$ and Σ. A different proof of a theorem illustrating the distributional and independence properties of these estimators will be presented. Unbiased estimators of $\underset{\sim}{\mu}$ and Σ will also be determined. Finally, the Wishart distribution and its properties will be presented.

1. The Maximum Likelihood Estimators of $\underset{\sim}{\mu}$ and Σ

The estimation problem for the multivariate case is very similar to the univariate problem. Let $X_1,\ldots,X_n$ be a random sample of size n from X, where $X \sim N(\mu,\sigma^2)$. Let $\hat{\mu}(X_1,\ldots,X_n)$ and $\hat{\sigma}^2(X_1,\ldots,X_n)$ denote the maximum likelihood estimators of μ and σ^2, respectively. Then, it is well-known that $\hat{\mu}(X_1,\ldots,X_n) = \bar{X}$ and $\hat{\sigma}^2(X_1,\ldots,X_n) = (1/n) \sum_{h=1}^{n} (X_h-\bar{X})^2$.

In the multivariate problem, the population from which the observation vectors are drawn is p-variate normal with density function given by Equation (1). Then for the p variables of interest it is possible to record n observation vectors $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n$. These observation vectors can be put in the form of an $(p \times n)$ data matrix X, with $n > p$, where

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n \end{bmatrix}.$$

Morrison (55) prefers to call the above matrix X the transpose of the data matrix. It is assumed that $f_{\underset{\sim}{X}_1', \ldots, \underset{\sim}{X}_n'}(\underset{\sim}{x}_1', \ldots, \underset{\sim}{x}_n') = f_{\underset{\sim}{X}_1'}(\underset{\sim}{x}_1') \ldots f_{\underset{\sim}{X}_n'}(\underset{\sim}{x}_n')$; that is, observation vectors were drawn independently of one another. Let $\hat{\underset{\sim}{\mu}}$ and $\hat{\Sigma}$ denote the maximum likelihood estimators of $\underset{\sim}{\mu}$ and Σ, respectively, with the understanding that both $\hat{\underset{\sim}{\mu}}$ and $\hat{\Sigma}$ are functions of $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$. Thus, $(\hat{\underset{\sim}{\mu}}, \hat{\Sigma})$ must be chosen so that for any admissible value $(\underset{\sim}{\mu}, \Sigma)$, $L(X|\hat{\underset{\sim}{\mu}}, \hat{\Sigma}) \geq L(X|\underset{\sim}{\mu}, \Sigma)$, where $L(X|\underset{\sim}{\mu}, \Sigma)$ denotes the likelihood function. $L(X|\underset{\sim}{\mu}, \Sigma)$ is usually written as L. The term likelihood stresses the interpretation that L is a function of unknown $\underset{\sim}{\mu}$ and Σ while the sample observations are fixed. To facilitate mathematical operations it is convenient to deal with ℓn L, the logarithmic likelihood function, since ℓn L assumes its maximum at the same point as does L. Since

$$L = (2\pi)^{-pn/2} |\Sigma|^{-n/2} \exp\left(-(1/2) \sum_{h=1}^{n} (\underset{\sim}{x}_h - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{x}_h - \underset{\sim}{\mu})\right), \tag{2}$$

$$\ell n\, L = -(1/2) pn\ \ell n (2\pi) + (n/2) \ell n |\Sigma^{-1}| - (1/2) \sum_{h=1}^{n} (\underset{\sim}{x}_h - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{x}_h - \underset{\sim}{\mu}). \tag{3}$$

The derivation of the maximum likelihood estimators of $\underset{\sim}{\mu}$ and Σ^{-1} as suggested by Anderson (3) and others will follow the presentation of two definitions and two lemmas.

Definition 3.1. The ($p\times 1$) sample mean vector, denoted by $\bar{\underset{\sim}{X}}$, has as its components the sample means of the p-variates. That is,

$$\bar{\underset{\sim}{X}} = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{bmatrix} = \begin{bmatrix} (1/n)\sum_{h=1}^{n} X_{1h} \\ \vdots \\ (1/n)\sum_{h=1}^{n} X_{ph} \end{bmatrix}.$$

Definition 3.2. The ($p\times p$) matrix of sums of squares and cross-products of deviations about the sample mean vector, denoted by A, is defined to be

$$A = \sum_{h=1}^{n} (\underset{\sim}{X}_h - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_h - \bar{\underset{\sim}{X}})'$$

$$= \begin{bmatrix} \sum_{h=1}^{n}(X_{1h}-\bar{X}_1)^2 & \cdots & \sum_{h=1}^{n}(X_{1h}-\bar{X}_1)(X_{ph}-\bar{X}_p) \\ & \vdots & \\ & \sum_{h=1}^{n}(X_{ih}-\bar{X}_i)(X_{jh}-\bar{X}_j) & \\ & \vdots & \\ \sum_{h=1}^{n}(X_{ph}-\bar{X}_p)(X_{1h}-\bar{X}_1) & \cdots & \sum_{h=1}^{n}(X_{ph}-\bar{X}_p)^2 \end{bmatrix}$$

A will also denote the matrix of observed values.

Lemma 3.1. If $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ are each ($p\times 1$) vectors, then for the ($p\times 1$) vector $\underset{\sim}{\mu}$,

$$\sum_{h=1}^{n}(\underset{\sim}{X}_h-\underset{\sim}{\mu})(\underset{\sim}{X}_h-\underset{\sim}{\mu})' = \sum_{h=1}^{n}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})' + n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})'.$$

Proof. Note that $(\underset{\sim}{X}_h-\underset{\sim}{\mu}) = (\underset{\sim}{X}_h-\bar{\underset{\sim}{X}}) + (\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})$. Then $\sum_{h=1}^{n}(\underset{\sim}{X}_h-\underset{\sim}{\mu})(\underset{\sim}{X}_h-\underset{\sim}{\mu})' = \sum_{h=1}^{n}\left((\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})+(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})\right)\left((\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})+(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})\right)'$. Note that for general matrices, B and C, $(B+C)' = B' + C'$. Then, application of the summation operator yields

$$\sum_{h=1}^{n}(\underset{\sim}{X}_h-\underset{\sim}{\mu})(\underset{\sim}{X}_h-\underset{\sim}{\mu})' = \sum_{h=1}^{n}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})' + \sum_{h=1}^{n}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})'$$
$$+ \sum_{h=1}^{n}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})' + \sum_{h=1}^{n}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})'.$$

But, $\sum_{h=1}^{n}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}}) = \sum_{h=1}^{n}\underset{\sim}{X}_h-n\bar{\underset{\sim}{X}} = n\left(n^{-1}\sum_{h=1}^{n}\underset{\sim}{X}_h\right) - n\bar{\underset{\sim}{X}} = n\bar{\underset{\sim}{X}} - n\bar{\underset{\sim}{X}} = \underset{\sim}{0}$, and the second and third terms drop out. Hence,

$$\sum_{h=1}^{n}(\underset{\sim}{X}_h-\underset{\sim}{\mu})(\underset{\sim}{X}_h-\underset{\sim}{\mu})' = \sum_{h=1}^{n}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})' + n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})'. \quad \|$$

Equivalently, $\sum_{h=1}^{n}(\underset{\sim}{X}_h-\underset{\sim}{\mu})(\underset{\sim}{X}_h-\underset{\sim}{\mu})' = A + n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})'$.

Lemma 3.2. Let B be a $(p\times p)$ matrix of independent real variables b_{ij}. Recall that the minor of b_{ij} is the determinant of the square submatrix of B obtained by deleting the *ith* row and *jth* column, and the cofactor of b_{ij}, denoted by B_{ij}, is the minor of b_{ij} multiplied by $(-1)^{i+j}$. Then

(i) $\partial|B|/\partial b_{ij} = B_{ij}$; and

(ii) if A is symmetric, then $\partial|A|/\partial a_{ii} = A_{ii}$ while $\partial|A|/\partial a_{ij} = 2A_{ij}$, $i \neq j$.

Proof. (i) It is well known that for $j = 1,2,\ldots,p$,

$$|B| = b_{1j}B_{1j} + \ldots + b_{ij}B_{ij} + \ldots + b_{pj}B_{pj}.$$

Since B_{ij} does not involve b_{ij}, then $\partial|B|/\partial b_{ij} = B_{ij}$.

(ii) Suppose that the $(p\times p)$ matrix B in (i) is such that $b_{ij} = f_{ij}(a_{11},\ldots,a_{pp})$, i.e., each element b_{ij} of B is a real-valued function of $(p\times p)$ independent real variables $a_{11},\ldots,a_{pp}$. Then

$$\partial|B|/\partial a_{rs} = \sum_{i=1}^{p}\sum_{j=1}^{p}(\partial|B|/\partial b_{ij})(\partial b_{ij}/\partial a_{rs}),$$

by the chain rule of partial derivatives. Expansion of the double summation yields

$$\begin{aligned}\frac{\partial|B|}{\partial a_{rs}} = {} & \frac{\partial|B|}{\partial b_{11}}\frac{\partial b_{11}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{1r}}\frac{\partial b_{1r}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{1s}}\frac{\partial b_{1s}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{1p}}\frac{\partial b_{1p}}{\partial a_{rs}} +\ldots \\ & + \frac{\partial|B|}{\partial b_{r1}}\frac{\partial b_{r1}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{rr}}\frac{\partial b_{rr}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{rs}}\frac{\partial b_{rs}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{rp}}\frac{\partial b_{rp}}{\partial a_{rs}} +\ldots \\ & + \frac{\partial|B|}{\partial b_{s1}}\frac{\partial b_{s1}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{sr}}\frac{\partial b_{sr}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{ss}}\frac{\partial b_{ss}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{sp}}\frac{\partial b_{sp}}{\partial a_{rs}} +\ldots \\ & + \frac{\partial|B|}{\partial b_{p1}}\frac{\partial b_{p1}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{pr}}\frac{\partial b_{pr}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{ps}}\frac{\partial b_{ps}}{\partial a_{rs}} +\ldots+ \frac{\partial|B|}{\partial b_{pp}}\frac{\partial b_{pp}}{\partial a_{rs}}.\end{aligned}$$

By hypothesis, $b_{ij} = f_{ij}(a_{11},\ldots,a_{pp})$. Suppose $b_{rs} = a_{rs}$ and $b_{sr} = a_{rs}$, $r,s = 1,\ldots,p$, $r \le s$. Then $\partial b_{rs}/\partial a_{rs} = 1$ and $\partial b_{sr}/\partial a_{rs} = 1$. From (i), $(\partial|B|/\partial b_{ij}) = B_{ij}$. Hence,

$$\partial|B|/\partial a_{rs} = (\partial|B|/\partial b_{rs}) + (\partial|B|/\partial b_{sr}) = B_{rs} + B_{sr}.$$

Now $|A| = |B|$ and $B_{rs} = B_{sr} = A_{rs} = A_{sr}$. Hence,

$$\partial|A|/\partial a_{rs} = 2A_{rs},$$

or

$$\partial|A|/\partial a_{ij} = 2A_{ij}.$$

From (i), it obviously follows that

$$\partial|A|/\partial a_{ii} = A_{ii}. \quad \|$$

<u>Theorem 3.1</u>. If $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ constitutes a random sample from a p-variate normal with parameters $\underset{\sim}{\mu}$ and Σ, then the maximum likelihood estimators of $\underset{\sim}{\mu}$ and Σ are $\hat{\underset{\sim}{\mu}}$ and $\hat{\Sigma}$, respectively, where

$$\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{X}} = (1/n) \sum_{h=1}^{n} \underset{\sim}{X}_h, \tag{4}$$

and

$$\hat{\Sigma} = (1/n) \sum_{h=1}^{n} (\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})' = (1/n)A \tag{5}$$

Proof. Recall that, in general, for two matrices B and C, tr BC = tr CB, where tr denotes the trace operator. Also, the trace of a scalar equals the scalar. Using these two properties on the last term of ℓn L, Equation (3), yields

$$\sum_{h=1}^{n} (\underset{\sim}{x}_h-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_h-\underset{\sim}{\mu}) = \text{tr} \sum_{h=1}^{n} (\underset{\sim}{x}_h-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_h-\underset{\sim}{\mu})$$

$$= \text{tr} \sum_{h=1}^{n} \Sigma^{-1}(\underset{\sim}{x}_h-\underset{\sim}{\mu})(\underset{\sim}{x}_h-\underset{\sim}{\mu})'$$

$$= \text{tr}\ \Sigma^{-1} \sum_{h=1}^{n} (\underset{\sim}{x}_h-\underset{\sim}{\mu})(\underset{\sim}{x}_h-\underset{\sim}{\mu})'.$$

Using Lemma 3.1 and noting that $\underset{\sim}{X}_h$ and $\bar{\underset{\sim}{X}}$ are now realizations,

$$\sum_{h=1}^{n} (\underset{\sim}{x}_h-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_h-\underset{\sim}{\mu}) = \text{tr}\ \Sigma^{-1}\left(A + n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\right)$$

$$= \text{tr}\ \Sigma^{-1}A + \text{tr}\ n\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'$$

$$= \text{tr}\ \Sigma^{-1}A + \text{tr}\ n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})$$

$$= \text{tr}\ \Sigma^{-1}A + n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}).$$

Then Equation (3) can be written as

$$\ell n\ L = -(1/2)pn\ \ell n(2\pi) + (n/2)\ell n|\Sigma^{-1}| - (1/2)\text{tr}\ \Sigma^{-1}A - (n/2)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}).$$

A necessary condition for the stationary values of ℓn L is given by

$$\partial \ell n\ L/\partial \underset{\sim}{\mu} = \underset{\sim}{0} \quad \text{and} \quad \partial \ell n\ L/\partial \Sigma^{-1} = 0.$$

Now,

$$\partial \ell n\ L/\partial \underset{\sim}{\mu} = n\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}).$$

Setting this equal to $\underset{\sim}{0}$ implies

$$(\bar{\underset{\sim}{X}}-\hat{\underset{\sim}{\mu}}) = \underset{\sim}{0}.$$

Hence, the maximum likelihood estimator of the parameter $\underset{\sim}{\mu}$ is

$$\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{X}}.$$

That $\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{X}}$ is obvious from examination of the last term of ℓn L which is $-(n/2)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})$. In Theorem 2.2, the matrix Σ^{-1} was shown to be positive definite. Hence $(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) > 0$ for $(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \neq 0$. And, this quadratic form is zero only when $\bar{\underset{\sim}{x}} = \underset{\sim}{\mu}$.

To find $\hat{\Sigma}^{-1}$, it is necessary to maximize the second and third terms of ℓn L which are

$$(n/2)\ell n|\Sigma^{-1}| - (1/2)\mathrm{tr}\ \Sigma^{-1}A.$$

The last term in ℓn L is zero since $\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{X}}$. The quantity, tr $\Sigma^{-1}A$, may be written as

$$\sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij} a_{ji},$$

where σ^{ij} is the ij*th* element of Σ^{-1} and a_{ji} is the ji*th* element of A. Thus, it is necessary to maximize

$$(n/2)\ln|\Sigma^{-1}| - (1/2) \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij} a_{ji}.$$

Maxima of these two terms are obtained by setting equal to zero the derivatives with respect to the elements of Σ^{-1}. Using Lemma 3.2,

$$\partial/\partial\sigma^{rr}\left((n/2)\ln|\Sigma^{-1}| - (1/2) \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij} a_{ji}\right) \tag{6}$$

$$= (n/2)|\Sigma^{-1}|^{-1}(\partial|\Sigma^{-1}|/\partial\sigma^{rr}) - (1/2)a_{rr}$$

$$= (n/2)|\Sigma^{-1}|^{-1}\mathrm{cof}\sigma^{rr} - (1/2)a_{rr},$$

where $\mathrm{cof}\sigma^{rr}$ denotes the cofactor of σ^{rr} in Σ^{-1}. For $r \neq s$,

$$(\partial/\partial\sigma^{rs})\left((n/2)\ln|\Sigma^{-1}| - (1/2) \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij} a_{ji}\right) = n|\Sigma^{-1}|^{-1}\mathrm{cof}\sigma^{rs} - a_{rs}. \tag{7}$$

In both Equation (6) and (7), there are expressions of the form $(\mathrm{cof}\sigma^{rs}|\Sigma^{-1}|^{-1})$. Now, $(\mathrm{cof}\sigma^{rs}|\Sigma^{-1}|^{-1})$ is the rs*th* element of Σ, which is denoted by σ_{rs}. Equations (6) and (7) are then set equal to zero to yield

$$n\hat{\sigma}_{rr} - a_{rr} = 0,$$

and

$$n\hat{\sigma}_{rs} - a_{rs} = 0,$$

for r,s = 1,...,p. In matrix form, these equations become

$$n\hat{\Sigma} = A,$$

or

$$\hat{\Sigma}^{-1} = nA^{-1},$$

assuming A^{-1} exists. Hence,

$$\hat{\Sigma}^{-1} = n\left\{\sum_{h=1}^{n}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})'\right\}^{-1},$$

assuming that A^{-1} exists. To find the maximum likelihood estimator of Σ, the invariant property of these estimators is used. Anderson (3) states this as follows:

> If on the basis of a given sample $\hat{\theta}_1,\ldots,\hat{\theta}_m$ are maximum likelihood estimates of the parameters $\theta_1,\ldots,\theta_m$ of a distribution, then $\phi_1(\hat{\theta}_1,\ldots,\hat{\theta}_m),\ldots,\phi_m(\hat{\theta}_1,\ldots,\hat{\theta}_m)$ are maximum likelihood estimates of $\phi_1(\theta_1,\ldots,\theta_m),\ldots,\phi_m(\theta_1,\ldots,\theta_m)$ if the transformation from $\theta_1,\ldots,\theta_m$ to $\phi_1,\ldots,\phi_m$ is one-to-one. If the estimates of $\theta_1,\ldots,\theta_m$ are unique, then the estimates of $\phi_1,\ldots,\phi_m$ are unique.

Then $\hat{\Sigma} = (1/n)A$. $\|$

When $p = 1$, $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = (1/n) \sum_{h=1}^{n} (X_h - \bar{X})^2$ which are the univariate results stated earlier.

2. Distributional Properties of $\hat{\mu}$ and $\hat{\Sigma}$

At this point, it would be desirable to find the distributions and certain other properties of $\hat{\mu}$ and $\hat{\Sigma}$. It will again be noted that the similarity between the univariate and multivariate results is very striking. Let $X_1, \ldots, X_n$ be a random sample from X, where $X \sim N(\mu, \sigma^2)$. Then, it can be shown that $\bar{X}$ and S_n^2 are independent random variables, where $\bar{X} = (1/n) \sum_{h=1}^{n} X_h$ and $S_n^2 = (n-1)^{-1} \sum_{h=1}^{n} (X_h - \bar{X})^2$. It can also be demonstrated that $\bar{X} \sim N(\mu, (\sigma^2/n))$ and $(n-1)S_n^2 = \sum_{h=1}^{n} (X_h - \bar{X})^2$ is distributed as $\sum_{h=2}^{n} Y_h^2$ where the Y_h are independent and $Y_h \sim N(0, \sigma^2)$. Finally, S_n^2 has been defined so that $E(S_n^2) = \sigma^2$.

The multivariate version of the above statements will be demonstrated using the concept of the direct product or Kronecker product of matrices. For alternative derivations, the reader should consult Anderson (3) or Roy (60).

The direct product concept has been used sparingly in the past. The design of experiments has been one limited area of application. The direct product concept is useful in showing the independence of $\bar{X}$ and S since it permits the viewpoint of $(np \times 1)$ spaces rather than $(p \times n)$ spaces. Only the definitions and lemmas necessary for the proof of Theorem 3.2 will be introduced. Graybill (26) and Lancaster (51) provide the reader with recent additional information concerning the direct product of matrices and statistical applications.

Definition 3.3. The (left) direct product of matrices A and B of sizes $(m_1 \times n_1)$ and $(m_2 \times n_2)$, respectively, is a matrix C, denoted $A \otimes B$, of size $(m_1 m_2 \times n_1 n_2)$ defined by

$$C = A \otimes B = \begin{bmatrix} Ab_{11} & Ab_{12} & \cdots & Ab_{1n_2} \\ \vdots & \vdots & \vdots & \vdots \\ Ab_{m_2 1} & Ab_{m_2 2} & \cdots & Ab_{m_2 n_2} \end{bmatrix} = [C_{ij}],$$

for $i = 1,\ldots,m_2$, and $j = 1,\ldots,n_2$. Note that $Ab_{ij} = b_{ij}A$.

Thus C is composed of $(m_2 \times n_2)$ submatrices, C_{ij}, where each submatrix has m_1 rows and n_1 columns. This definition will be illustrated with the following examples.

Example 3.1. Let

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad B = [4,5].$$

Then,

$$C = A \otimes B = [4A, 5A] = \left[\begin{array}{cc|cc} 4 & 8 & 5 & 10 \\ 0 & 12 & 0 & 15 \end{array}\right],$$

where $C_{11} = [4A] = \begin{bmatrix} 4 & 8 \\ 0 & 12 \end{bmatrix}$ and $C_{12} = [5A] = \begin{bmatrix} 5 & 10 \\ 0 & 15 \end{bmatrix}$.

Example 3.2. Let A be any $(m_1 \times n_1)$ matrix and let I be the $(m_2 \times m_2)$ identity matrix. Then $A \otimes I$ is the $(m_1 m_2 \times m_2 n_1)$ block diagonal matrix whose diagonal entries are A. That is,

$$A \otimes I = \begin{bmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A \end{bmatrix}.$$

If the $(m_1 \times n_1)$ matrix A and $(m_2 \times n_2)$ matrix B have entries $a_{\alpha\beta}$ and b_{ij}, respectively, then the element of $A \otimes B$ at the intersection of the row, $(i-1)m_2 + \alpha$, and of the column, $(j-1)n_2 + \beta$, is $a_{\alpha\beta} b_{ij}$. In Example 3.1, the element of $A \otimes B$ at the intersection of the row, $(1-1)2 + 2 = 2$, and of the column, $(2-1)2 + 2 = 4$, is $a_{22} b_{12} = 15$.

Lemma 3.3. Let A be an $(m_1 \times n_1)$ matrix, B an $(m_2 \times n_2)$ matrix, C an $(n_1 \times k_1)$ matrix, and D an $(n_2 \times k_2)$ matrix. Then

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

Proof. $A \otimes B$ is of the form $[b_{\alpha\beta} A]$ where $\alpha = 1, \ldots, m_2$ and $\beta = 1, \ldots, n_2$. $C \otimes D$ is of the form $[C d_{\beta\gamma}]$ where $\beta = 1, \ldots, n_2$ and $\gamma = 1, \ldots, k_2$. If we apply the rule for multiplying partitioned matrices, we obtain a matrix of the form, $[ACf_{\alpha\gamma}]$, where $f_{\alpha\gamma} = \sum_{\beta=1}^{n_2} b_{\alpha\beta} d_{\beta\gamma}$. Thus $f_{\alpha\gamma}$ is nothing more than the $\alpha\gamma th$ element of the ordinary matrix product BD. Hence, $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ and has size $(m_1 m_2 \times k_1 k_2)$. ||

A useful result of the above lemma is that for any $(m_1 \times n_1)$ matrix A and any $(m_2 \times n_2)$ matrix B it follows that $A \otimes B = (A \otimes I_{n_2})(I_{n_1} \otimes B)$.

Lemma 3.4. If A is an $(m \times m)$ nonsingular matrix and B is an $(n \times n)$ nonsingular matrix, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

Proof. $(A \otimes B)(A^{-1} \otimes B^{-1}) = AA^{-1} \otimes BB^{-1} = I_m \otimes I_n = I_{mn}$. ||

The above lemma is sometimes stated as the inverse of a direct product is the direct product of the inverses.

Lemma 3.5. For any two matrices, A and B, $(A \otimes B)' = A' \otimes B'$.

Proof. $A \otimes B = [Ab_{ij}]$. So $(A \otimes B)' = [A'b_{ij}'] = [A'b_{ji}]$. But, $A' \otimes B' = [A'b_{ji}]$. Thus $(A \otimes B)' = A' \otimes B'$. ||

This lemma is sometimes stated as the transpose of the direct product is the direct product of the transposes.

Aitken (2) stated the following, which facilitates the proof of Theorem 3.2:

> Let X be an arbitrary rectangular matrix of order $m \times n$. Let us suppose its elements written down, row after row, as the mn ordered elements of a vector

$$\underset{\sim}{\xi}' = (x_{11}, x_{12}, \ldots, x_{1n}, x_{21}, x_{22}, \ldots, x_{2n}, \ldots, x_{mn}).$$

> Now let $Y = H'XK$, where H and K may be rectangular, and let a vector $\underset{\sim}{\eta}$ be written down for Y in the same way as $\underset{\sim}{\xi}\ldots$, we see that $\underset{\sim}{\eta} = (K'\otimes H)\underset{\sim}{\xi}$, the transforming matrix being the direct product.

To see this, we inspect y_{ij}. Let $Y = [y_{ij}]$, $H = [h_{id}]$, $X = [x_{de}]$, and $K = [k_{ej}]$. Then the dj*th* element of XK is $\sum_e x_{de}k_{ej}$, and $y_{ij} = \sum_d \sum_e h_{di}x_{de}k_{ej}$. Now $K'\otimes H$ is of the form $[K'h_{id}]$. And, from examination of the coefficient of x_{de}, $Y = H'XK$ is identical to $\underset{\sim}{\eta} = (K'\otimes H)\underset{\sim}{\xi}$.

Using the preceding lemmas and Aitken's result on the direct product, another viewpoint of some of the earlier results of this chapter may be obtained.

Let $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ be a random sample from $\underset{\sim}{X}$, where $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma_p)$. Then the data matrix X, $n > p$, is

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1n} \\ & \vdots & \\ X_{p1} & \cdots & X_{pn} \end{bmatrix} = [\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n].$$

By the definition of a random sample, the $\underset{\sim}{X}_h$ are independent for all h. Also, $EX = [\underset{\sim}{\mu}, \ldots, \underset{\sim}{\mu}] = \underset{\sim}{\mu}\underset{\sim}{j}_n'$, where $\underset{\sim}{j}_n$ is the $(n\times 1)$ vector $[1,1,\ldots,1]'$. But $\underset{\sim}{\mu}\underset{\sim}{j}_n' = I_p\underset{\sim}{\mu}\underset{\sim}{j}_n'$. Let dictionary form denote that a $(p\times n)$ matrix has its elements written down, row after row, as the pn ordered elements of a vector. Let $\underset{\sim}{\xi}$ denote X written in dictionary form. Then $\underset{\sim}{\mu}_{\underset{\sim}{\xi}} = (\underset{\sim}{j}_n\otimes I_p)\underset{\sim}{\mu}$, using Aitken's result. Also, $C(\underset{\sim}{\xi}, \underset{\sim}{\xi}') = \Sigma_{\underset{\sim}{\xi}} = I_n\otimes\Sigma_p$. Thus, the $(np\times 1)$ vector $\underset{\sim}{\xi} \sim N\left((\underset{\sim}{j}_n\otimes I_p)\underset{\sim}{\mu},\ I_n\otimes\Sigma_p\right)$. It is now possible to prove Theorem 3.2.

<u>Theorem 3.2.</u> Let $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ be a random sample from a p-variate normal ($n>p$) where $\underset{\sim}{X}\sim N(\underset{\sim}{\mu},\Sigma_p)$. Then

(i) $\bar{\underset{\sim}{X}}\sim N(\underset{\sim}{\mu},(1/n)\Sigma_p)$;

(ii) $n\hat{\Sigma}$ is distributed as $\sum_{h=2}^{n} \underset{\sim}{Y}_h\underset{\sim}{Y}_h'$ where $\underset{\sim}{Y}_h\sim N(\underset{\sim}{0},\Sigma_p)$ and the $\underset{\sim}{Y}_h$ are independent for $h = 1,\ldots,n$;

(iii) $\bar{\underset{\sim}{X}}$ and $\hat{\Sigma}$ are independent.

Proof. Consider the transformation

$$Y = I_p X H_n',$$

where H_n is an ($n\times n$) orthogonal matrix and the first row of H_n is $[1/\sqrt{n},\ldots,1/\sqrt{n}]$. Let $\underset{\sim}{\eta}$ denote Y written in dictionary form. That is

$$\underset{\sim}{\eta} = [Y_{11},\ldots,Y_{1n},Y_{21},\ldots,Y_{2n},\ldots,Y_{pn}]'.$$

The use of Aitken's result gives

$$\underset{\sim}{\eta} = (H_n \otimes I_p)\underset{\sim}{\xi},$$

and

$$\underset{\sim}{\mu}_{\underset{\sim}{\eta}} = (H_n \otimes I_p)\underset{\sim}{\mu}_{\underset{\sim}{\xi}} = (H_n\underset{\sim}{j}_n \otimes I_p)\underset{\sim}{\mu},$$

where $H_n\underset{\sim}{j}_n = \left[\sqrt{n},\ \sum_{j=1}^{n} h_{2j},\ldots,\ \sum_{j=1}^{n} h_{nj}\right]' = [\sqrt{n},0,\ldots,0]'.$

Also,

$$C(\underset{\sim}{\eta},\underset{\sim}{\eta}') = \Sigma_{\underset{\sim}{\eta}} = (H_n \otimes I_p)(I_n \otimes \Sigma_p)(H_n \otimes I_p)'$$

$$= (I_n \otimes \Sigma_p).$$

Hence,

$$\underset{\sim}{\eta} \sim N\left((H_n \underset{\sim}{j}_n \otimes I_p)\underset{\sim}{\mu},\ I_n \otimes \Sigma_p\right).$$

In the transformation, $Y = I_p X H_n'$, the orthogonal matrix H_n can be written as

$$H_n = \begin{bmatrix} \underset{\sim}{h}_1' \\ \underset{\sim}{h}_2' \\ \vdots \\ \underset{\sim}{h}_n' \end{bmatrix} = \begin{bmatrix} 1/\sqrt{n} & 1/\sqrt{n} & \cdots & 1/\sqrt{n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix}$$

while

$$H_n' = [\underset{\sim}{h}_1, \underset{\sim}{h}_2, \ldots, \underset{\sim}{h}_n].$$

Then

$$\underset{\sim}{Y}_1 = (\underset{\sim}{h}_1' \otimes I_p)\underset{\sim}{\xi} = \sqrt{n}\ \bar{\underset{\sim}{X}},$$

where $\underset{\sim}{Y}_1 = [Y_{11},\ldots,Y_{p1}]'$. Since $\underset{\sim}{Y}_1 = (\underset{\sim}{h}_1'\otimes I_p)\underset{\sim}{\xi}$ and $\underset{\sim}{\xi}\sim N(\underset{\sim}{\mu}_{\underset{\sim}{\xi}},\Sigma_{\underset{\sim}{\xi}})$, then $\underset{\sim}{Y}_1\sim N\big((\underset{\sim}{h}_1'\otimes I_p)\underset{\sim}{\mu}_{\underset{\sim}{\xi}},(h_1'\otimes I_p)\Sigma_{\underset{\sim}{\xi}}(h_1'\otimes I_p)'\big)$, which simplifies to $\underset{\sim}{Y}_1\sim N(\sqrt{n}\ \underset{\sim}{\mu},\Sigma_p)$. But $\underset{\sim}{Y}_1 = \sqrt{n}\ \bar{\underset{\sim}{X}}$, or $\bar{\underset{\sim}{X}} = n^{-1/2}\underset{\sim}{Y}_1$. Hence,

$$\bar{\underset{\sim}{X}}\sim N\big(\underset{\sim}{\mu},(1/n)\Sigma_p\big).$$

Consider the properties of $\underset{\sim}{Y}_h$, $h = 2,\ldots,n$. Let I_n be partitioned as follows:

$$I_n = \begin{bmatrix} \underset{\sim}{e}_1' \\ \underset{\sim}{e}_2' \\ \vdots \\ \underset{\sim}{e}_n' \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Thus $\underset{\sim}{e}_h$ has a 1 in the h*th* position and 0's elsewhere. Also, let $Y = [\underset{\sim}{Y}_1,\underset{\sim}{Y}_2,\ldots,\underset{\sim}{Y}_n]$. In view of this, $\underset{\sim}{Y}_h = (\underset{\sim}{e}_h'\otimes I_p)\underset{\sim}{\eta}$. The transformation, $(\underset{\sim}{e}_h'\otimes I_p)$, has the effect of selecting every n*th* element from $\underset{\sim}{\eta}$, starting with Y_{1h} and ending with Y_{ph}. Thus, for $h = 2,\ldots,n$,

$$\underset{\sim}{\mu}_{\underset{\sim}{Y}_h} = (\underset{\sim}{e}_h'\otimes I_p)\underset{\sim}{\mu}_{\underset{\sim}{\eta}} = (\underset{\sim}{e}_h'\otimes I_p)(H_n\underset{\sim}{j}_n\otimes I_p)\underset{\sim}{\mu} = (0\otimes I_p)\underset{\sim}{\mu} = \underset{\sim}{0},$$

and

$$\Sigma_{\underset{\sim}{Y}_h} = (\underset{\sim}{e}_h'\otimes I_p)\Sigma_{\underset{\sim}{\eta}}(\underset{\sim}{e}_h'\otimes I_p)' = (\underset{\sim}{e}_h'\otimes I_p)(I_n\otimes\Sigma_p)(\underset{\sim}{e}_h\otimes I_p)$$

$$= (\underset{\sim}{e}_h'\underset{\sim}{e}_h \otimes \Sigma_p) = (1 \otimes \Sigma_p) = \Sigma_p.$$

Hence, for $h = 2,\ldots,n$, $\underset{\sim}{Y}_h \sim N(\underset{\sim}{0},\Sigma_p)$. To show that $\underset{\sim}{Y}_h$ are independent for all h, examine the covariance matrix of $\underset{\sim}{Y}_i$ and $\underset{\sim}{Y}_j$ for all i and j. Since Y has an np-variate normal, $\underset{\sim}{Y}_i$ and $\underset{\sim}{Y}_j$ are jointly independent if and only if, for all $i \neq j$, the covariance matrix is the 0 matrix.

$$C(\underset{\sim}{Y}_i,\underset{\sim}{Y}_j') = (\underset{\sim}{e}_i'\otimes I_p)C(\underset{\sim}{n},\underset{\sim}{n}')(\underset{\sim}{e}_j'\otimes I_p)'$$

$$= (\underset{\sim}{e}_i'\otimes I_p)(I_n\otimes\Sigma_p)(\underset{\sim}{e}_j\otimes I_p)$$

$$= (\underset{\sim}{e}_i'\otimes\Sigma_p)(\underset{\sim}{e}_j\otimes I_p) = \delta_{ij}\otimes\Sigma_p,$$

where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$. Therefore,

$$C(\underset{\sim}{Y}_i,\underset{\sim}{Y}_j') = 0, \quad \text{if } i \neq j$$

$$= \Sigma_p, \text{ if } i = j.$$

And the $\underset{\sim}{Y}_h$ are independent for all h. From this, it may be stated that $\bar{\underset{\sim}{X}}$ and $\hat{\Sigma}$ are independent. First, note that

$$n\hat{\Sigma} = \sum_{h=1}^{n}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})' = \sum_{h=1}^{n}\underset{\sim}{X}_h\underset{\sim}{X}_h' - n\bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}'$$

$$= XX' - n\bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}'$$

$$= XH_n'H_nX' - (\sqrt{n}\ \underset{\sim}{\bar{X}})(\sqrt{n}\ \underset{\sim}{\bar{X}})'$$

$$= YY' - \underset{\sim}{Y}_1\underset{\sim}{Y}_1' = \sum_{h=2}^{n} \underset{\sim}{Y}_h\underset{\sim}{Y}_h'.$$

Thus, $n\hat{\Sigma}$ is distributed as $\sum_{h=2}^{n} \underset{\sim}{Y}_h\underset{\sim}{Y}_h'$ where $\underset{\sim}{Y}_h \sim N(\underset{\sim}{0},\Sigma_p)$. Finally, since $\underset{\sim}{Y}_h$ are independent for all h and $\underset{\sim}{\bar{X}}$ is a function only of $\underset{\sim}{Y}_1$ and $\hat{\Sigma}$ is not a function of $\underset{\sim}{Y}_1$, we see that $\underset{\sim}{\bar{X}}$ and $\hat{\Sigma}$ are independent. ||

In the preceding theorem, it was demonstrated that $E\underset{\sim}{\hat{\mu}} = E\underset{\sim}{\bar{X}} = \underset{\sim}{\mu}$. Thus, $\underset{\sim}{\hat{\mu}}$ is an unbiased estimator of $\underset{\sim}{\mu}$. But

$$E\hat{\Sigma} = E(1/n)A = (1/n)E \sum_{h=2}^{n} \underset{\sim}{Y}_h\underset{\sim}{Y}_h' = (1/n)(n-1)\Sigma_p.$$

To overcome the bias, let $S = (n-1)^{-1}A$. Then

$$ES = (n-1)^{-1}EA = (n-1)^{-1}(n-1)\Sigma_p = \Sigma_p.$$

Thus, S is an unbiased estimator of Σ_p, and $\underset{\sim}{\bar{X}}$ and S will be used as the estimators for $\underset{\sim}{\mu}$ and Σ, respectively.

3. Properties of the Wishart Distribution

It was also demonstrated that $n\hat{\Sigma} = \sum_{h=1}^{n} (\underset{\sim}{X}_h-\underset{\sim}{\bar{X}})(\underset{\sim}{X}_h-\underset{\sim}{\bar{X}})' = \sum_{h=2}^{n} \underset{\sim}{Y}_h\underset{\sim}{Y}_h'$, where $\underset{\sim}{Y}_h \sim N(\underset{\sim}{0},\Sigma)$ and the $\underset{\sim}{Y}_h$ are independent. In the sequel, the distribution of $A = \sum_{h=1}^{\nu} \underset{\sim}{Y}_h\underset{\sim}{Y}_h'$, where $\underset{\sim}{Y}_h \sim N(\underset{\sim}{0},\Sigma)$ and $\underset{\sim}{Y}_h$ are independent, will be needed. Various methods have been presented for the derivation of this distribution. Based on a random sample of n p-variate vectors from a p-variate normal population, Wishart (72) transformed the np-fold sample normal by introducing quadratic coordinates and integrated out the undesired variables. A review of the other methods has been presented by Wishart (73). Anderson (3) also presents a detailed derivation, as does Kshirsager (49). Suppose the ($p\times 1$) vectors $\underset{\sim}{Y}_h$ are independent and $\underset{\sim}{Y}_h \sim N(\underset{\sim}{0},\Sigma)$, $h = 1,2,\ldots,\nu$, then the density of $A = \sum_{h=1}^{\nu} \underset{\sim}{Y}_h \underset{\sim}{Y}_h'$ is given by

$$\frac{|A|^{(\nu-p-1)/2}\, e^{-(1/2)\mathrm{tr}\Sigma^{-1}A}}{2^{\nu p/2}\pi^{p(p-1)/4}|\Sigma|^{\nu/2}\prod_{h=1}^{p}\Gamma\big((\nu+1-h)/2\big)}, \tag{8}$$

for A positive definite and 0 otherwise. It is customary to denote this density by $w(A|\Sigma,p,\nu)$. Since $n\hat{\Sigma} = \sum_{h=1}^{n-1} \underset{\sim}{Y}_h \underset{\sim}{Y}_h'$, an immediate consequence is that $A = n\hat{\Sigma}$ has density $w(A|\Sigma,p,n-1)$. The Wishart also possesses the reproductive property. That is, if A_h has density $w(A_h|\Sigma,p,\nu_h)$, $h = 1,2,\ldots,k$, then $A = \sum_{h=1}^{k} A_h$ has density $w(A|\Sigma,p,\sum_{h=1}^{k}\nu_h)$.

The following two lemmas are introduced here since they deal with the Wishart. However, they will not be used until Chapter V. These lemmas were suggested by Bowker (7) and permit a concise derivation of Hotelling's T^2 distribution under general conditions.

Lemma 3.6. Let the elements of the ($p\times p$) symmetric random matrix A have density function $f(A)$. Let the elements of the ($p\times p$) orthogonal, symmetric random matrix B be distributed independently of the elements of A with density function $g(B)$. Also, let H be a ($p\times p$) orthogonal matrix such that $C = H'AH$. Then, if the function f has the property that

$f(A) = f(C)$, the matrix $A^* = BAB'$ is a $(p\times p)$ symmetric random matrix whose elements have density function $f(A^*)$ and are distributed independently of B.

Proof. The independence of the elements of A and B imply that the joint density of A and B is $f(A)g(B)$. Let $A^* = BAB'$. If $\underset{\sim}{a}^*$ and $\underset{\sim}{a}$ denote A^* and A, respectively, written in dictionary form, then the transformations $A^* = BAB'$ and $\underset{\sim}{a}^* = (B\otimes B')\underset{\sim}{a}$ are the same, and the Jacobian of the transformation is $|B'\otimes B| = |B|^p|B'|^p = |BB'|^p = 1$. Since $A^* = BAB'$, then $A = B'A^*B$ and the joint density of A^* and B is $f(B'A^*B)g(B)$. But f has the property that $f(A) = f(A^*)$. Thus, $f(B'A^*B)g(B) = f(A^*)g(B)$, and the elements of A^* are distributed independently of the elements of B. ||

Note that $w(A|I,p,\nu)$ satisfies the property that $f(A) = f(C)$, where $C = H'AH$. That is, $w(A|I,p,\nu) = w(C|I,p,\nu)$, where $C = H'AH$.

Lemma 3.7. Let the elements of the $(p\times p)$ symmetric random matrix A have density function $w(A|I,p,\nu)$. Let the $(p\times p)$ matrix A and its inverse, A^{-1}, be partitioned as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{bmatrix},\quad A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{12'} & A^{22} \end{bmatrix},$$

where

$$A_{11} = [a_{ij}],\quad A^{11} = [a^{ij}],\quad i = 1,2,\ldots,q,\quad j = 1,2,\ldots,q.$$

Then $(A^{11})^{-1}$ has density function $w\left((A^{11})^{-1}|I_q,q,\nu-p+q\right)$.

Proof. Let $\left(c(p,\nu)\right)^{-1} = 2^{\nu p/2}\pi^{p(p-1)/4}|I|^{\nu/2}\prod_{i=1}^{p}\Gamma\left((\nu+1-i)/2\right)$, and write $w(A|I,p,\nu)$ as $w(A_{11},A_{12},A_{22}|I,p,\nu)$ in view of the partitioning of A. Then

$$w(A_{11},A_{12},A_{22}|I,p,\nu)$$

$$= c(p,\nu)|A|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}\ A\right)$$

$$= c(p,\nu)(|A_{22}||A_{11}-A_{12}A_{22}^{-1}A_{12}'|)^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}\ A_{11}-(1/2)\mathrm{tr}\ A_{22}\right)$$

$$= c(p,\nu)|A_{11}-A_{12}A_{22}^{-1}A_{12}'|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}\ A_{11}\right)$$

$$\cdot\ |A_{22}|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}\ A_{22}\right),$$

by Lemma 2.5 and the properties of the tr operator. Make the transformation from (A_{11},A_{12},A_{22}) to $\left((A^{11})^{-1},A_{12},A_{22}\right)$, where $(A^{11})^{-1} = A_{11}-A_{12}A_{22}^{-1}A_{12}'$ by Lemma 2.4. Then,

$$w\left((A^{11})^{-1},A_{12},A_{22}|I,p,\nu\right)$$

$$= c(p,\nu)|(A^{11})^{-1}|^{(\nu-p-1)/2}\exp\left\{-(1/2)\mathrm{tr}\left((A^{11})^{-1} + A_{12}A_{22}^{-1}A_{12}'\right)\right\}$$

$$\cdot\ |A_{22}|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}\ A_{22}\right)$$

$$= c(p,\nu)\left|(A^{11})^{-1}\right|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}(A^{11})^{-1}\right)$$

$$\cdot\left|A_{22}\right|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}(A_{22}+A_{12}A_{22}^{-1}A_{12}')\right)$$

$$= c(q,\nu-p+q)\left|(A^{11})^{-1}\right|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}(A^{11})^{-1}\right)$$

$$\cdot\left(c(p,\nu)/c(q,\nu-p+q)\right)\left|A_{22}\right|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}(A_{22}+A_{12}A_{22}^{-1}A_{12}')\right),$$

where $c(q,\nu-p+q)\left|(A^{11})^{-1}\right|^{(\nu-p-1)/2}\exp\left(-(1/2)\mathrm{tr}(A^{11})^{-1}\right) =$ $w\left((A^{11})^{-1}\middle| I_q,q,\nu-p+q\right)$. ‖

Let $X_1,X_2,\ldots,X_n$ be a random sample of size n from X, where $X\sim N(\mu,\sigma^2)$. Then $\sum_{h=1}^{n} (X_h-\bar{X})^2/\sigma^2\sim\chi^2(n-1)$. If $p = 1$ with $A = \sum_{h=1}^{n} (X_h-\bar{X})^2$ and $\Sigma = \sigma^2$, then the Wishart density of Equation (8) reduces to that of a chi-square with n - 1 degrees of freedom. This and other properties of the Wishart will be further utilized in Chapters V and VII.

Chapters II and III provide the scenario for the sequel. The properties of $\underset{\sim}{\bar{X}}$ and S presented in Chapter III are not new. However, the salient aspect of this chapter is the direct product proof of Theorem 3.2. This form of proof does not seem to have previously appeared in the literature.

CHAPTER IV

THEORETICAL CONTROL CHARTS FOR THE MEAN

A brief introduction to statistical quality control via control charts was presented in Chapter I. This chapter will consider the use of control charts for maintaining surveillance of the mean of p quality characteristics for a process which is in an existing state of statistical control and the parameters of the distribution from which the sample is drawn are *known*.

1. Control Charts for the Mean When Σ is Known

Suppose p quality characteristics are of interest and are governed by a p-variate normal distribution with known parameters $\underset{\sim}{\mu}_0$ and Σ. It is desired that the process remain at the nominal value $\underset{\sim}{\mu}_0$. Since the manufacturing process could shift out of control at any time, one desires that a p-variate control chart detect this with a reasonably high probability. To determine whether a process mean is in control at a given time, a random sample of size n is obtained and a realization of some statistic is determined from this sample data. When $p = 1$, $\bar{X}$ is the statistic used. For this reason and since $\bar{\underset{\sim}{X}} \sim N(\underset{\sim}{\mu}_0, (1/n)\Sigma)$ for a random sample of size n from $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}_0, \Sigma)$, it seems reasonable that $\bar{\underset{\sim}{X}}$ should be used to maintain surveillance on $\underset{\sim}{\mu}_0$ for $p > 1$. The value of the parameter $\underset{\sim}{\mu}_0$ may be derived from past data. If this is the case, the amount of data on which $\underset{\sim}{\mu}_0$ is based is to be considered large enough such that μ_0 may be treated as a value of the parameter and not its estimate. Duncan (18) states that μ_0 could have also been selected by management to attain certain objectives. However, management must be careful in selecting μ_0 since the process may not be in control at μ_0 but at some other level. The values of the p(p+1)/2 different entries of Σ could have also been derived from past data or selected by management in accordance with a specific policy.

For one quality characteristic, the control chart for the sample mean is denoted by three horizontal lines: an upper control limit (UCL), a central line (CL or ℄), and a lower control limit (LCL). To maintain control over μ_0 by using $\bar{X}$, a k-sigma control chart is constructed with control limits given by

$$\mathrm{UCL} = \mu_0 + (k/\sqrt{n})\sigma$$

$$\mathrm{CL} = \mu_0$$

$$\mathrm{LCL} = \mu_0 - (k/\sqrt{n})\sigma,$$

where k is usually taken to be 2 or 3. The values of the sample means are plotted on the chart. For k = 3, the control chart is illustrated in Figure 3. Some disturbances may cause the process mean to shift by an amount ε such that now $X \sim N\left(\mu_0+\varepsilon,(1/n)\sigma^2\right)$. Under the new state, the probability that the first sample value of $\bar{X}$ detects this is given by

$$1 - P\left(\mu_0-k(\sigma/\sqrt{n})\le\bar{X}\le\mu_0+k(\sigma/\sqrt{n})\,\middle|\,\mu_{\bar{X}}=\mu_0+\varepsilon\right),$$

where the vertical bar indicates a condition on $\mu_{\bar{X}}$. This probability equals

$$1 - \left[\Phi\left(k - \frac{\varepsilon}{(\sigma/\sqrt{n})}\right) - \Phi\left(-k - \frac{\varepsilon}{(\sigma/\sqrt{n})}\right)\right],$$

where Φ denotes the distribution function of the univariate standard normal random variable. The greater the departure of $\varepsilon/(\sigma/\sqrt{n})$ from zero, the greater the probability that the control chart will detect this shift. Instead of control charts, the decision maker could resort to the less visual but equivalent technique of repeated tests of significance.

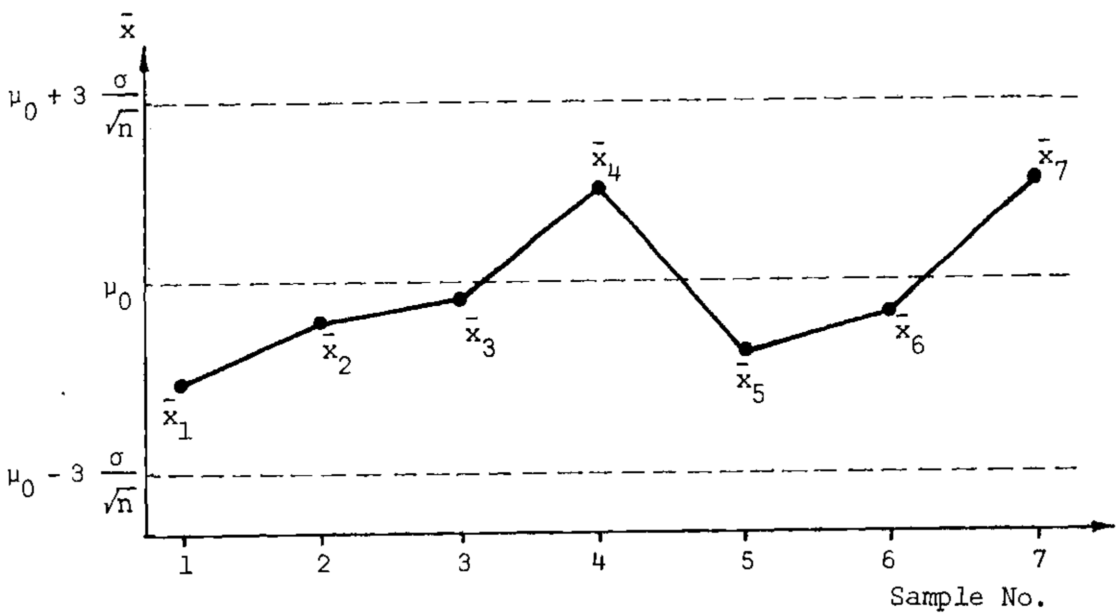

Figure 3. A Univariate Three Sigma $\bar{x}$ Chart

To denote that the p-variate normal density function depends directly on the parameters $\underset{\sim}{\mu}$ and Σ, some authors prefer to write $f_{\underset{\sim}{X}'}(\underset{\sim}{x}')$ as $f_{\underset{\sim}{X}'}(\underset{\sim}{x}';\underset{\sim}{\mu},\Sigma)$. The admissible values of $\underset{\sim}{\mu}$ and Σ determine the parameter space Ω, which is a subset of $[p+p(p+1)/2]$-dimensional Euclidean space. Let Ω be partitioned into the two nonempty subsets: ω and $\Omega - \omega$, where

$$\omega = \{(\underset{\sim}{\mu},\Sigma)\colon \underset{\sim}{\mu}=\underset{\sim}{\mu}_0,\ \Sigma \text{ is fixed}\},$$

$$\Omega - \omega = \{(\underset{\sim}{\mu},\Sigma)\colon \underset{\sim}{\mu}\neq\underset{\sim}{\mu}_0,\ \Sigma \text{ is fixed}\}.$$

To determine whether the mean vector has shifted from the nominal value $\underset{\sim}{\mu}_0$, set up the null hypothesis

$$H_0\colon \{f_{\underset{\sim}{X}_1',\ldots,\underset{\sim}{X}_n'}(\underset{\sim}{x}_1',\ldots,\underset{\sim}{x}_n';\underset{\sim}{\mu},\Sigma)\colon (\underset{\sim}{\mu},\Sigma)\epsilon\omega\}$$

against the alternative hypothesis

$$H_1\colon \{f_{\underset{\sim}{X}_1',\ldots,\underset{\sim}{X}_n'}(\underset{\sim}{x}_1',\ldots,\underset{\sim}{x}_n';\underset{\sim}{\mu},\Sigma)\colon (\underset{\sim}{\mu},\Sigma)\epsilon\Omega-\omega\},$$

where the symbol ϵ denotes membership in the designated set. These hypotheses are usually written

$$H_0\colon \underset{\sim}{\mu} = \underset{\sim}{\mu}_0 \quad \text{versus} \quad H_1\colon \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0.$$

Let W denote the ($p \times n$) dimensional sample space and w a subset of W such that, on the basis of the data matrix X, the decision rule will have the form:

$$\text{Do not reject } H_0 \text{ if } X \in W\text{-}w$$

$$\text{Reject } H_0 \text{ if } X \in w,$$

where w is usually called the critical region or rejection region for H_0. The probability of incorrectly rejecting the null hypothesis, denoted by α, is called the significance level or size of the test. That is,

$$\alpha = P(X \in w \mid \underset{\sim}{\mu} = \underset{\sim}{\mu}_0), \tag{9}$$

where the vertical bar indicates that the probability statement is conditional upon the truth of the null hypothesis. The form of the critical region will be determined by using the likelihood-ratio principle. Thus, one looks at

$$\lambda(X) = \frac{\sup\{L(X \mid \underset{\sim}{\mu}, \Sigma) : (\underset{\sim}{\mu}, \Sigma) \in \omega\}}{\sup\{L(X \mid \underset{\sim}{\mu}, \Sigma) : (\underset{\sim}{\mu}, \Sigma) \in \Omega\}} . \tag{10}$$

When $p = 1$, the test of H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$ where σ is known yields

$$\lambda(\underset{\sim}{x}') = \frac{\sup\{L(\underset{\sim}{x}'|\mu,\sigma^2):\ \mu=\mu_0\}}{\sup\{L(\underset{\sim}{x}'|\mu,\sigma^2):\ -\infty<\mu<\infty\}} ,$$

where $L(\underset{\sim}{x}'|\mu,\sigma^2) = (2\pi)^{-n/2}\sigma^{-n}\exp\left(-(1/2\sigma^2)\sum_{h=1}^{n}(x_h-\mu)^2\right)$. In the denominator of $\lambda(\underset{\sim}{x}')$, the sup is attained by using the maximum likelihood estimator of μ. Thus,

$$\lambda(\underset{\sim}{x}') = \frac{(2\pi)^{-n/2}\sigma^{-n}\exp\left(-(1/2\sigma^2)\sum_{h=1}^{n}(x_h-\mu_0)^2\right)}{(2\pi)^{-n/2}\sigma^{-n}\exp\left(-(1/2\sigma^2)\sum_{h=1}^{n}(x_h-\bar{x})^2\right)}$$

$$= \exp\left(-(1/2\sigma^2)n(\bar{x}-\mu_0)^2\right),$$

and the form of the critical region is given by

$$w = \{\underset{\sim}{x}':\ \left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right)^2 > c\} . \tag{11}$$

Equivalently,

$$w = \{\underset{\sim}{x}':\ \mathrm{mod}\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| > c_1\},$$

where c_1 is usually chosen so that

$$P\left(\text{mod}\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right)>c_1 \middle| H_0 \text{ is true}\right) = \alpha.$$

This distributional problem also needs to be solved. Since $\bar{X}\sim N\left(\mu_0,(\sigma^2/n)\right)$, then $\left(\sqrt{n}\ (\bar{X}-\mu_0)/\sigma\right) \sim N(0,1)$, and c_1 should satisfy

$$P\left(-c_1 \le \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \le c_1\right) = 1 - \alpha.$$

For $1 - \alpha = .9973$, $c_1 = 3$. Thus, when the decision maker establishes 3-sigma control limits and takes a random sample of size n, he is actually testing H_0: $\mu = \mu_0$ versus H_1: $\mu \ne \mu_0$ with $\alpha = .0027$ and $w = \{\underset{\sim}{x}'$: $\text{mod}\left(\sqrt{n}(\bar{x}-\mu_0)/\sigma\right)>3.0\}$.

Let $\chi^2(p)$ and $\chi^2(p,\alpha)$ denote the chi-square random variable with p degrees of freedom and the upper α-percentile point of the chi-square distribution with p degrees of freedom (d.f.), respectively. Since $\left(\sqrt{n}(\bar{X}-\mu_0)/\sigma\right)\sim N(0,1)$, then $\left(\sqrt{n}(\bar{X}-\mu_0)/\sigma\right)^2\sim\chi^2(1)$. If the decision maker decides to use the form of the critical region given by Equation (11), his distributional problem would consist of finding a constant c such that

$$P\left(\chi^2(1)>c \middle| \mu=\mu_0\right) = \alpha.$$

For $\alpha = .0027$, $c = \chi^2(1,.0027) = 9.0$. Since $\chi^2(1)$ is a generalized measure of distance, the control chart would now appear as in Figure 4 with only an upper control limit. Similar results will be obtained for the multivariate problem. The center line of the univariate χ^2 control chart is the horizontal axis since when a sample mean equals μ_0 the calculated value of the χ^2 statistic is zero. The simplicity of construction of the χ^2 chart is offset somewhat by the fact that runs above and below the mean will be harder to detect since they are intermingled.

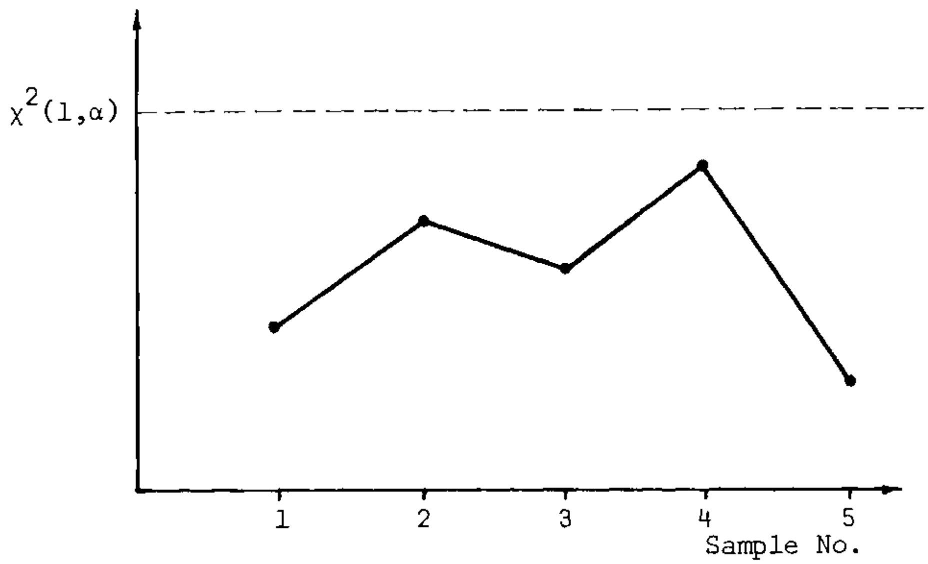

Figure 4. A Univariate Chi-Square Control Chart

To test H_0: $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ versus H_1: $\underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ by the likelihood ratio principle, Equation (10) becomes

$$\lambda(X) = \frac{(2\pi)^{-pn/2}|\Sigma|^{-n/2}\exp\left(-(1/2)\sum_{h=1}^{n}(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)\right)}{(2\pi)^{-pn/2}|\Sigma|^{-n/2}\exp\left(-(1/2)\sum_{h=1}^{n}(\underset{\sim}{x}_h-\bar{\underset{\sim}{x}})'\Sigma^{-1}(\underset{\sim}{x}_h-\bar{\underset{\sim}{x}})\right)},$$

where the sup of the denominator was attained by using Equation (4) of Theorem 3.1. As stated in the proof of Theorem 3.1,

$$\sum_{h=1}^{n}(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0) = \operatorname{tr}\,\Sigma^{-1}\sum_{h=1}^{n}(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)'.$$

Similarly,

$$\begin{aligned}\sum_{h=1}^{n}(\underset{\sim}{x}_h-\bar{\underset{\sim}{x}})'\Sigma^{-1}(\underset{\sim}{x}_h-\bar{\underset{\sim}{x}}) &= \operatorname{tr}\,\Sigma^{-1}\sum_{h=1}^{n}(\underset{\sim}{x}_h-\bar{\underset{\sim}{x}})(\underset{\sim}{x}_h-\bar{\underset{\sim}{x}})' \\ &= \operatorname{tr}\,\Sigma^{-1}\left[\sum_{h=1}^{n}(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)'-n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\right],\end{aligned}$$

by the use of Lemma 3.1. Thus,

$$\lambda(X) = \exp\left(-(1/2)\operatorname{tr}\,\Sigma^{-1}n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\right) = \exp\left(-(1/2)n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)\right),$$

and the form of the critical region is given by

$$w = \{X:\ n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)>c\}. \tag{12}$$

The general distributional problem involves the noncentral chi-square distribution.

<u>Definition 4.1</u>. If the $(p\times 1)$ vector $\underset{\sim}{X}\sim N(\underset{\sim}{\mu},I)$, then $\underset{\sim}{X}'\underset{\sim}{X}$ is distributed as a noncentral chi-square random variable with p degrees of freedom and noncentrality parameter $\lambda = \underset{\sim}{\mu}'\underset{\sim}{\mu}$, denoted by $\chi'^2(p,\lambda)$. $\chi'^2(p,\lambda,\alpha)$ will

denote the upper α percentile point of $\chi'^2(p,\lambda)$.

There have been numerous derivations of the noncentral chi-square distribution. The following presentation is based on van der Vaart's derivation (69). If $X_i \sim N(\mu_i,1)$, then

$$M_{X_i^2}(t) = (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} \exp\left(tx_i^2-(x_i-\mu_i)^2/2\right)dx_i = (1-2t)^{-1/2}\exp\left((\mu_i^2 t)/(1-2t)\right).$$

Since the X_i are independent, the X_i^2 are independent for $i = 1,\ldots,p$, and

$$M_{\chi'^2(p,\lambda)}(t) = \prod_{h=1}^{p} (1-2t)^{-1/2}\exp\left((\mu_h^2 t)/(1-2t)\right)$$

$$= (1-2t)^{-p/2}\exp\left((\lambda t)/(1-2t)\right).$$

To obtain the corresponding noncentral chi-square density function note that

$$t/(1-2t) = (1/2)\left((2t)/(1-2t)\right) \quad \text{and} \quad (2t)/(1-2t) = -1 + \left(1/(1-2t)\right).$$

Thus,

$$M_{\chi'^2(p,\lambda)}(t) = (1-2t)^{-p/2}\exp\left((\lambda/2)(-1+(1-2t)^{-1})\right).$$

The expansion of $\exp\left(\lambda/2(1-2t)\right)$ in its Maclaurin series yields

$$M_{\chi'^2(p,\lambda)}(t) = (1-2t)^{-p/2}\exp(-\lambda/2) \sum_{h=0}^{\infty} \left(\lambda/2(1-2t)\right)^h (1/h!)$$

$$= \sum_{h=0}^{\infty} (1-2t)^{-(1/2)(p+2h)} \frac{\exp(-\lambda/2)(\lambda/2)^h}{h!} .$$

Since the moment-generating function is a linear operator on density functions, it follows that the density function of $\chi'^2(p,\lambda)$ is

$$f_{\chi'^2(p,\lambda)}(s) = \sum_{h=0}^{\infty} \frac{\exp(-\lambda/2)(\lambda/2)^h}{h!} f_{\chi^2(p+2h)}(s),$$

where $f_{\chi^2(p+2h)}$ is the density function of a central chi-square random variable with $(p+2h)$d.f. Thus $f_{\chi'^2(p,\lambda)}(s)$ may be regarded as a weighted mixture of central χ^2 density functions where the weights are Poisson probabilities. By successive differentiation of the moment-generating function, it immediately follows that $E\chi'^2(p,\lambda) = p + \lambda$, and $V\chi'^2(p,\lambda) = 2p + 4\lambda$. Also note that when $\lambda = 0$, $\chi'^2(p,\lambda)$ reduces to $\chi^2(p)$. A solution to the general distributional problem can now be formulated.

<u>Theorem 4.1</u>. If $\bar{\underset{\sim}{X}}\sim N\left(\underset{\sim}{\mu},(1/n)\Sigma\right)$, then $n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)\sim\chi'^2(p,\lambda)$ where $\lambda = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$.

Proof. Since Σ is positive definite, there exists a nonsingular matrix R such that $\Sigma = RR'$ (see Definition 2.1). Let $Z = \sqrt{n}\, R^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)$. Then, $E\underset{\sim}{Z} = \sqrt{n}\, R^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$ and $C(\underset{\sim}{Z},\underset{\sim}{Z}') = E(\underset{\sim}{Z}-\underset{\sim}{\mu}_{\underset{\sim}{Z}})(\underset{\sim}{Z}-\underset{\sim}{\mu}_{\underset{\sim}{Z}})' = nR^{-1}E(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})'(R^{-1})' = R^{-1}\Sigma(R^{-1})' = I$. Thus, $\underset{\sim}{Z}\sim N\left(\sqrt{n}\, R^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0),I\right)$, and $\underset{\sim}{Z}'\underset{\sim}{Z}\sim\chi'^2(p,\lambda)$ where $\lambda = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'(R^{-1})'R^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0) = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'(RR')^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0) = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$.

But $\underset{\sim}{Z}'\underset{\sim}{Z} = n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'(R^{-1})'R^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0) = n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)$. Hence, $n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0) \sim \chi'^2(p,\lambda)$, with $\lambda = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$. ||

When $\bar{\underset{\sim}{X}} \sim N(\underset{\sim}{\mu}_0,(1/n)\Sigma)$, $\lambda = n(\underset{\sim}{\mu}_0-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}_0-\underset{\sim}{\mu}_0) = 0$ and $n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0) \sim \chi^2(p)$. Hence, the critical constant c in Equation (12) is the percentage point $\chi^2(p,\alpha)$ such that

$$P\left(n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0) > \chi^2(p,\alpha) \mid \underset{\sim}{\mu} = \underset{\sim}{\mu}_0\right) = \alpha.$$

In testing H_0: $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ versus H_1: $\underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$, the decision maker would take a random sample of size n, compute $\bar{\underset{\sim}{x}}$, and determine whether $n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0) > \chi^2(p,\alpha)$, where $\chi^2(p,\alpha)$ is obtained from tables of the chi-square distribution. Thus the decision maker sets up a chart similar to the one in Figure 4 where the control limit $\chi^2(1,\alpha)$ is replaced by $\chi^2(p,\alpha)$. A warning line could have also been inserted at $\chi^2(p,\alpha')$ where $\alpha < \alpha'$, $0 < \alpha < 1$. When the decision maker computes $n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)$ and compares it with the control limit for successive samples of size n, he is merely performing repeated tests of significance. The control charts are called theoretical because the limits have been determined without the use of any information in the current sample. The construction of the limits for the χ^2 control chart may use levels of α for which the percentage points are not easily accessible from the standard statistics texts. A nomogram has been prepared by Boyd (11) which permits a rapid determination of these points. Johnson and Kotz (45) report on an approximation by Hill (31) which is supposedly quite accurate for ε near to 0 or 1. This approximation is given by

$$\chi^2(p,\varepsilon) \approx \left(p-(2/3)\right)\exp[z_\varepsilon c-(1/6)z_\varepsilon^2+(1/36)(z_\varepsilon^3-z_\varepsilon)c^{-1}$$

$$-(1/1620)(6z_\varepsilon^4-31z_\varepsilon^2-32)c^{-2}+(1/38880)(9z_\varepsilon^5-308z_\varepsilon^3-481z_\varepsilon)c^{-3}],$$

where $P\left(\chi^2(p)<\chi^2(p,\varepsilon)\right) = \varepsilon$, $P(Z<z_\varepsilon) = \varepsilon$, and $c = \left((1/2)p-(1/3)\right)^{1/2}$. Unfortunately, this approximation seems to be in error as can be easily seen for $\varepsilon = 0.9973$.

The hypothesis of control might be false; yet, on the basis of the observations, the decision maker might accept H_0 as true. This error will be referred to as type II error as opposed to type I error. The probability of type II error will be denoted by β. For the data matrix X,

$$\beta(\lambda) = P(X\epsilon W-w|\underset{\sim}{\mu}\neq\underset{\sim}{\mu}_0),$$

where $\beta(\lambda)$ denotes that this probability depends on the noncentrality parameter $\lambda = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$. The power of the test, denoted by π, is such that

$$\pi(\lambda) = P(X\epsilon w|\underset{\sim}{\mu}).$$

More specifically,

$$\pi(\lambda) = P\left[n(\underset{\sim}{\bar{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\bar{X}}-\underset{\sim}{\mu}_0)>\chi^2(p,\alpha)|\underset{\sim}{\mu}\right],$$

where the vertical bar indicates that the true density has $\underset{\sim}{\mu}$ as its population mean. For a continuum of parameter values, the resulting probabilities constitute the power function. When $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$, $\pi(\lambda=0) = \alpha$.

The distribution of $\chi'^2(p,\lambda)$ has been tabulated by Fix(22) for $\alpha = .01, .05$, $\pi = .1(.1).9$, and d.f. = 1(1)20, 22(2)40, 45(5)55, 60(10)100, where the numbers in parentheses indicate the step size. A latter and more extensive table, designed to facilitate power calculations, has been tabulated by Haynam et al. (29). Harter and Owen (28) have made this table more accessible with their publication. Haynam's Table I lists entries of the power for α = .001, .005, .01, .025, .05, .1, λ = 0(.1)1(.2)3.0(.5)5(1)40(2)50(5)100, and d.f. = 1(1)30(2)50(5)100. The univariate Central Limit Theorem also permits a simple approximation to the power; that is,

$$1 - \pi(\lambda) = F_{\chi'^2(p,\lambda)}\left(\chi^2(p,\alpha)\right) \approx \Phi\left(\frac{\chi^2(p,\alpha)-(p+\lambda)}{\sqrt{2(p+2\lambda)}}\right).$$

Charts of the power function for the noncentral F distribution have been prepared by Pearson and Hartley (56). These charts can be used for the power of a noncentral chi-square random variable and are discussed in Chapter V.

2. Determination of Out of Control Characteristics

When H_0: $\underset{\sim}{\mu}=\underset{\sim}{\mu}_0$ is rejected, the determination of those components of $\underset{\sim}{\mu}$ responsible for this rejection is of prime importance. For p = 2, an elliptical control chart could be of some use. This is discussed in

Jackson (40). However, for $p \geq 3$ or for a χ^2-control chart (Figure 4), the use of simultaneous techniques is now advocated. When the null hypothesis is rejected and the decision maker is faced with the nebulous alternative of not H_0, certain principles of simultaneous statistical inference will enable him to draw conclusions regarding which means are significant.

Definition 4.2. Let $\underset{\sim}{X}' = (X_1, X_2, \ldots, X_p)$ be jointly distributed random variables, whose joint distribution depends on the unknown parameters, $\underset{\sim}{\mu}' = (\mu_1, \mu_2, \ldots, \mu_p)$. For $h = 1, \ldots, p$, let $\underline{\mu}_h$ and $\bar{\mu}_h$ be 2p functions of the elements of the sample values of $\underset{\sim}{X}$ and let A_h be the event that the interval $\underline{\mu}_h$ to $\bar{\mu}_h$ covers μ_h. Then, a set of simultaneous confidence intervals for $\mu_1, \mu_2, \ldots, \mu_p$ consists of these 2p functions of the sample values with the property that

$$P\left(\bigcap_{h=1}^{p} A_h\right) = P(\underline{\mu}_1 < \mu_1 < \bar{\mu}_1, \ldots, \underline{\mu}_p < \mu_p < \bar{\mu}_p) \geq 1-\alpha.$$

If the inequality sign holds, the set is of bounded confidence level; that is, the probability of coverage is greater than or equal to $1 - \alpha$ rather than strictly equal to $1 - \alpha$. Only bounded confidence level sets will be considered.

Bonferroni Z Intervals

Two simultaneous techniques will be considered. The first of these will be called the Bonferroni Z technique since the derivation is based on Bonferroni's inequality which is based on Boole's inequality.

<u>Lemma 4.1 (Boole's Inequality).</u> For any p events $A_1, A_2, \ldots, A_p$,

$$P\left(\bigcup_{h=1}^{p} A_h\right) \leq \sum_{h=1}^{p} P(A_h).$$

Proof. If $A_1, A_2, \ldots, A_p$ is any collection of p events and $B_h = A_h \cap A_{h-1}^c \cap A_{h-2}^c \cap \ldots \cap A_1^c$, where A_i^c denotes the complement of A_i, then $A_1 \cup A_2 \cup \ldots \cup A_p = B_1 \cup B_2 \cup \ldots \cup B_p$ where the B_h are disjoint and $B_h \subset A_h$ for $h = 1,2,\ldots,p$. Clearly, $P(A_1 \cup A_2 \cup \ldots \cup A_p) = P(B_1) + P(B_2) + \ldots + P(B_p)$. Now $B_h \subset A_h$, so $P(B_h) \leq P(A_h)$. Hence, $P(A_1 \cup A_2 \cup \ldots \cup A_p) \leq P(A_1) + P(A_2) + \ldots + P(A_p)$.

The Bonferroni Z technique also uses the concept of a confidence interval for the unknown parameter μ of a normally distributed univariate random variable with known population variance σ^2. For a random sample of size n, the two statistics $\underline{\mu} = \bar{X} - z_{\gamma/2}(\sigma/\sqrt{n})$ and $\bar{\mu} = \bar{X} + z_{\gamma/2}(\sigma/\sqrt{n})$ constitute a $100(1-\gamma)\%$ confidence interval for μ, where $P(Z>z_{\gamma/2}) = (\gamma/2)$. Since $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma)$ and $X_h \sim N(\mu_h, \sigma_h^2)$, $h = 1,\ldots,p$, the aforementioned univariate results apply in constructing confidence intervals for each μ_h. The Bonferroni technique follows.

<u>Theorem 4.2.</u> Let $\underline{\mu}_h = \bar{X}_h - z_{\alpha/2p}(\sigma_h/\sqrt{n})$, $\bar{\mu}_h = \bar{X}_h + z_{\alpha/2p}(\sigma_h/\sqrt{n})$, and $(\underline{\mu}_h, \bar{\mu}_h)$ be $100\left(1-(\alpha/p)\right)\%$ confidence intervals for μ_h, where $\bar{X}_h = (1/n) \sum_{j=1}^{n} X_{hj}$, $h = 1,\ldots,p$. If A_h denotes the event that $(\underline{\mu}_h, \bar{\mu}_h)$ covers μ_h, then $P\left(\bigcap_{h=1}^{p} A_h\right) \geq 1 - \sum_{h=1}^{p} P(A_h^c) = 1-\alpha$.

Proof. For $h = 1,\ldots,p$, $P(A_h) = 1 - (\alpha/p)$, and $P(A_h^c) = (\alpha/p)$. The probability that all p events occur simultaneously is $P(A_1 \cap A_2 \cap \ldots \cap A_p)$. But, $P(A_1 \cap A_2 \cap \ldots \cap A_p) = 1 - P(A_1 \cap A_2 \cap \ldots \cap A_p)^c = 1 - P(A_1^c \cup A_2^c \cup \ldots \cup A_p^c)$, by De Morgan's Laws. By Lemma 4.1, $P(A_1^c \cup A_2^c \cup \ldots \cup A_p^c) \leq P(A_1^c) + P(A_2^c) + \ldots + P(A_p^c)$. Hence, $P(A_1 \cap A_2 \cap \ldots \cap A_p) \geq 1 - \left(P(A_1^c) + P(A_2^c) + \ldots + P(A_p^c)\right)$. That is, $P(A_1 \cap A_2 \cap \ldots \cap A_p) \geq 1 - p(\alpha/p) = 1-\alpha$. ‖

The statement of the conclusion for this theorem could also be written as

$$P\left\{\bigcap_{h=1}^{p}\left\{\operatorname{mod}\left(\frac{\bar{X}_h - \mu_h}{\sigma_h/\sqrt{n}}\right) \leq z_{\alpha/2p}\right\}\right\} \geq 1 - \alpha.$$

It should also be noted that the proof of this theorem does not require that $X_1, X_2, \ldots, X_p$ be independent random variables.

For each component interval above, the confidence coefficient was set at $1 - (\alpha/p)$. However, the equal confidence coefficients can be abandoned, and any unequal allocation can be substituted provided $\sum_{h=1}^{p} \alpha_h = \alpha$.

Table 1 has as its entries selected values of $1 - (\alpha/p)$, the confidence coefficient of the p individual confidence intervals. It is seen that, as p increases, the confidence coefficient of each of the p intervals also increases but at a decreasing rate.

Table 1. Values of $1 - (\alpha/p)$

$1-\alpha$ \ p	1	2	3	4	5
.95	.9500	.9750	.9833	.9875	.9900
.99	.9900	.9950	.9967	.9975	.9980

To compute the values of $\underline{\mu}_h$ and $\bar{\mu}_h$, the upper $(\alpha/2p)$ percentage points of the standard normal distribution are required. Table 2 has entries of $z_{\alpha/2p}$ for $p = 1(1)5$ and $\alpha = .01$ and $.05$.

Table 2. Values of $z_{\alpha/2p}$

$1-\alpha$ \ p	1	2	3	4	5
.95	1.96	2.24	2.395	2.495	2.575
.99	2.575	2.81	2.93	3.01	3.09

Scheffe's χ^2 Intervals

The second simultaneous technique to be considered will be called Scheffe's χ^2 technique or χ^2 projections since the derivation is based on methods proposed by Scheffe (62) in his F technique. Dunn (19,20) has given a brief discussion of this and other methods. Miller (53) has recently presented an overview of the general simultaneous statistical inference problem and includes a synopsis of Dunn's earlier works. Since Miller devotes most of his attention to Scheffe's F technique, a detailed development of the χ^2 technique will be presented. To obtain confidence intervals on the component means μ_h, $h = 1,2,\ldots,p$, simultaneous confidence intervals will be developed for the linear combinations $\underset{\sim}{a}'\underset{\sim}{\mu}$ for all nonnull $\underset{\sim}{a}$. These confidence intervals will in turn be used to obtain the confidence intervals for the coordinate means. Scheffe's χ^2 technique requires four lemmas.

Lemma 4.2. The test of the hypothesis H_0: $\underset{\sim}{\mu}=\underset{\sim}{\mu}_0$ is equivalent to testing the hypothesis H_0': $\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\underset{\sim}{\mu}_0$ for all nonnull $\underset{\sim}{a}$.

Proof. Suppose one is testing H_0': $\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\underset{\sim}{\mu}_0$ for all nonnull $\underset{\sim}{a}$. If $\underset{\sim}{e}_h' = (0,0,\ldots,1,\ldots,0)$, then H_0': $\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\underset{\sim}{\mu}_0$ becomes H_0: $\underset{\sim}{\mu}=\underset{\sim}{\mu}_0$ by successively letting $\underset{\sim}{a}' = \underset{\sim}{e}_h'$, $h = 1,\ldots,p$. Suppose one is testing H_0: $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$. For any given $\underset{\sim}{a}$, $\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\underset{\sim}{\mu}_0$ and the converse holds. $\|$

The proof of Lemma 4.3 depends on the Cauchy-Schwarz inequality which asserts that mod $(\underset{\sim}{a}'\underset{\sim}{b}) \le \|\underset{\sim}{a}\|\|\underset{\sim}{b}\|$ for any $\underset{\sim}{a}$ and $\underset{\sim}{b}$ in R^p, where $\|a\| = \left(\sum_{h=1}^{p} a_h^2\right)^{1/2}$.

Lemma 4.3. For $c > 0$, mod$(\underset{\sim}{a}'\bar{\underset{\sim}{x}}-\underset{\sim}{a}'\underset{\sim}{\mu}) \le c\|\underset{\sim}{a}\|$ for all $\underset{\sim}{a}$ if and only if $\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^2 \le c^2$.

Proof. Suppose $\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^2 \le c^2$. By the Cauchy-Schwarz inequality, mod$(\underset{\sim}{a}'\bar{\underset{\sim}{x}}-\underset{\sim}{a}'\underset{\sim}{\mu}) \le \|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|\|\underset{\sim}{a}\|$. Thus, mod$(\underset{\sim}{a}'\bar{\underset{\sim}{x}}-\underset{\sim}{a}'\underset{\sim}{\mu}) \le c\|\underset{\sim}{a}\|$. Suppose mod$(\underset{\sim}{a}'\bar{\underset{\sim}{x}}-\underset{\sim}{a}'\underset{\sim}{\mu}) \le c\|\underset{\sim}{a}\|$ for all $\underset{\sim}{a}$. Let $\underset{\sim}{a} = \bar{\underset{\sim}{x}} - \underset{\sim}{\mu}$. Then, $\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^4 \le c^2\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^2$. If $\bar{\underset{\sim}{x}} = \underset{\sim}{\mu}$, the inequality $\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^2 \le c^2$ is obvious. If $\bar{\underset{\sim}{x}} \ne \underset{\sim}{\mu}$, $\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^2$ may be cancelled, yielding $\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^2 \le c^2$. $\|$

The geometrical interpretation of the inequality $\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^2 \le c^2$ is that $\underset{\sim}{\mu}$ lies in the sphere of radius c centered at $\bar{\underset{\sim}{x}}$. If $\|\underset{\sim}{a}\|^2 = 1$, then the geometrical interpretation of $\mathrm{mod}(\underset{\sim}{a}'\bar{\underset{\sim}{x}}-\underset{\sim}{a}'\underset{\sim}{\mu}) \le c\|\underset{\sim}{a}\|$ is that $\underset{\sim}{\mu}$ lies between the two parallel planes tangent to the sphere $\|\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}\|^2 \le c^2$ and orthogonal to the vector $\underset{\sim}{a}$. Thus the lemma could be restated by saying that a closed sphere is equal to the intersection of all regions formed by parallel tangent planes. Lemma 4.4 is an extension of the previous lemma.

<u>Lemma 4.4</u>. For $c > 0$, $\mathrm{mod}(\underset{\sim}{a}'\bar{\underset{\sim}{x}}-\underset{\sim}{a}'\underset{\sim}{\mu}) \le c(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2}$ for all $\underset{\sim}{a}$ if and only if $(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \le c^2$, where $\bar{\underset{\sim}{X}} \sim N(\underset{\sim}{\mu},\Sigma/n)$.

Proof. Suppose $(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \le c^2$. From Definition 2.1, it is seen that there exists a nonsingular matrix R such that $\Sigma = RR'$. Let $\underset{\sim}{y} = R^{-1}\bar{\underset{\sim}{x}}$. Then the inequality $(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \le c^2$ becomes $(\underset{\sim}{y}-R^{-1}\underset{\sim}{\mu})'(\underset{\sim}{y}-R^{-1}\underset{\sim}{\mu}) \le c^2$, or $\|\underset{\sim}{y}-R^{-1}\underset{\sim}{\mu}\|^2 \le c^2$. By Lemma 4.3, $\mathrm{mod}(\underset{\sim}{\ell}'\underset{\sim}{y}-\underset{\sim}{\ell}'R^{-1}\underset{\sim}{\mu}) \le c\|\underset{\sim}{\ell}\|$ for all $\underset{\sim}{\ell}$. Let $\underset{\sim}{a} = (R^{-1})'\underset{\sim}{\ell}$. Since R is nonsingular, there is a one-to-one correspondence between $\underset{\sim}{\ell}$ and $\underset{\sim}{a}$. In terms of $\underset{\sim}{a}$, the above inequality becomes $\mathrm{mod}(\underset{\sim}{a}'\bar{\underset{\sim}{x}}-\underset{\sim}{a}'\underset{\sim}{\mu}) \le c(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2}$ for all $\underset{\sim}{a}$. The converse follows immediately. ||

The geometrical interpretation of Lemma 4.4 is very similar to that of Lemma 4.3 with ellipsoid substituted for circle. Thus, a closed ellipsoid is equal to the intersection of all regions formed by parallel tangent planes.

A plane of support to an ellipsoid may be defined as a plane that has at least one point in common with the ellipsoid and such that the ellipsoid is entirely on one side of the plane. Scheffe (63) presents a very thorough geometrical discussion of ellipsoids and their planes of support. Lemma 4.5 states the equations of two parallel planes of support to an ellipsoid which are orthogonal to a vector $\underset{\sim}{a}$. Miller (53) proves this for the case when the ellipsoid is a sphere, and Wilks (71) earlier gives this as an exercise.

Lemma 4.5. The equations of the two hyperplanes orthogonal to the vector $\underset{\sim}{a}$ and tangent to the ellipsoid $(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \leq c^2$, positioned at $\bar{\underset{\sim}{x}}$, are $\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\bar{\underset{\sim}{x}} \pm c(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2}$.

Proof. For fixed $\underset{\sim}{a}$, the equation $\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{k}$ defines a (p-1) dimensional hyperplane in the space of $\underset{\sim}{\mu}$. If $\underset{\sim}{a}^* = (k\underset{\sim}{a})/\|\underset{\sim}{a}\|^2$, then $\underset{\sim}{a}'\underset{\sim}{a}^* = k$ and $\underset{\sim}{a}^*$ lies in this plane. For any $\underset{\sim}{\mu}$ in the plane, $(\underset{\sim}{\mu}-\underset{\sim}{a}^*)'\underset{\sim}{a}^* = 0$, which shows that the plane $\underset{\sim}{a}'\underset{\sim}{\mu} = k$ is orthogonal to $\underset{\sim}{a}^*$ and, thus, to $\underset{\sim}{a}$. Hence, the planes $\underset{\sim}{a}'\underset{\sim}{\mu} = k$, with $k = \underset{\sim}{a}'\bar{\underset{\sim}{x}} \pm c\|\underset{\sim}{a}\|$ are also orthogonal to $\underset{\sim}{a}$. To establish the tangency of the planes to the ellipsoid, minimize the function $(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})$ subject to the constraint $\underset{\sim}{a}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \pm c(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2} = 0$. Differentiation of the lagrangian function $f(\underset{\sim}{\mu},\lambda) = (\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) + \lambda[\underset{\sim}{a}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})\pm c(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2}]$ yields $-2\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) - \lambda\underset{\sim}{a} = \underset{\sim}{0}$ and $\underset{\sim}{a}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \pm c(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2} = 0$. Simultaneous solution yields $\underset{\sim}{\mu} = \bar{\underset{\sim}{x}} \pm [c/(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2}]\Sigma\underset{\sim}{a}$, which are the points at which the planes actually touch the ellipsoid. Finally, $\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\bar{\underset{\sim}{x}} \pm c(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2}$. ‖

From Equation (12), it is seen that the acceptance region for testing H_0: $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ versus H_1: $\underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ is given by $\{\bar{\underset{\sim}{x}}: n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0) \leq \chi^2(p,\alpha)\}$, which is an ellipsoid in p-dimensional space centered at

μ_0. Lemma 4.2 stated that the test of H_0: $\mu = \mu_0$ is equivalent to the test H_0': $a'\mu = a'\mu_0$ for all nonnull a. Thus, the acceptance region for this test is also given by $\{\bar{x}: n(\bar{x}-\mu_0)'\Sigma^{-1}(\bar{x}-\mu_0) \le \chi^2(p,\alpha)\}$. Equivalently, one could look at the ellipsoidal *confidence* region centered at $\bar{x}$ and specified by $\{\mu: n(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu) \le \chi^2(p,\alpha)\}$. The relation between the confidence regions for μ and $a'\mu$ for all nonnull a is the basis for the next theorem.

Theorem 4.3. Let $X_1, X_2, \ldots, X_n$ be a random sample from X, where $X \sim N(\mu, \Sigma)$ with μ unknown and Σ known. Then

$$P[(a'(\bar{X}-\mu))^2 \le n^{-1}\chi^2(p,\alpha)(a'\Sigma a) \text{ for all nonnull } a] = 1 - \alpha. \quad (13)$$

Proof. From Equation (12), it is seen that $P[n(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu) \le \chi^2(p,\alpha)]$ $= P[(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu) \le n^{-1}\chi^2(p,\alpha)] = 1 - \alpha$. From Lemma 4.4, it follows that $\mathrm{mod}(a'\bar{x}-a'\mu) \le (n^{-1}\chi^2(p,\alpha))^{1/2}(a'\Sigma a)^{1/2}$ for all a if and only if $(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu) \le n^{-1}\chi^2(p,\alpha)$. Hence, $P[\mathrm{mod}(a'\bar{X}-a'\mu) \le$ $(n^{-1}\chi^2(p,\alpha))^{1/2}(a'\Sigma a)^{1/2}$ for all nonnull $a] = 1 - \alpha$, and Equation (13) follows. ||

With probability $(1-\alpha)$, a point μ is contained in the ellipsoisal confidence region if and only if it lies between *all* pairs of supporting hyperplanes orthogonal to a for all nonnull a. An immediate extension of Equation (13) and Lemma 4.5 is that the equations of these hyperplanes are given by

$$\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\bar{\underset{\sim}{x}} \pm \left(n^{-1}\chi^2(p,\alpha)\right)^{1/2}(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2}. \tag{14}$$

To obtain confidence intervals for the coordinate means, μ_h, $h = 1,\ldots,$ p, let $\underset{\sim}{a}' = \underset{\sim}{e}_h'$. Then confidence interval estimates for $\underset{\sim}{e}_h'\underset{\sim}{\mu} = \mu_h$ are given by

$$\bar{x}_h \pm \left(n^{-1}\chi^2(p,\alpha)\right)^{1/2}(\sigma_{hh})^{1/2}, \tag{15}$$

where σ_{hh} is the h*th* diagonal element of Σ. In developing these confidence interval estimates for the coordinate means only a finite number of the vectors $\underset{\sim}{a}$ were used. Upon examination of Equation (13), it is seen that the probability that the p statements, $\mu_n \epsilon \bar{x}_h \pm \left(n^{-1}\chi^2(p,\alpha)\right)^{1/2}(\sigma_{hh})^{1/2}$, are simultaneously true is greater than $1 - \alpha$. Figure 5 below illustrates this concept for $p = 2$.

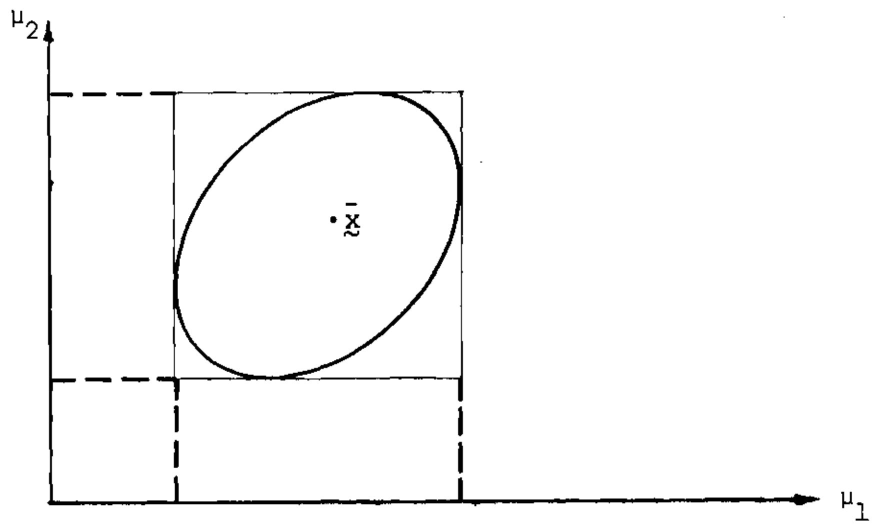

Figure 5. Scheffe Projections for p = 2

Comparison of Intervals

Of prime interest to the practicing quality control decision maker is the question of which intervals are better: the Bonferroni intervals or Scheffe's intervals. Suppose it is desired to determine those coordinate means which led to the rejection of H_0: $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$. From Theorem 4.2, it is seen that the Bonferroni interval estimates are given by

$$\bar{x}_h \pm z_{\alpha/2p} n^{-1/2} (\sigma_{hh})^{1/2} \tag{16}$$

Thus, the only difference between the Bonferroni and Scheffe intervals are the constants $z_{\alpha/2p}$ and $\left(\chi^2(p,\alpha)\right)^{1/2}$, respectively. If the criterion of the shortest confidence interval is adopted, it would be useful to have some idea of which of these percentage points is smaller. Now $P(Z \geq z_{\alpha/2p}) = (\alpha/2p)$ is equivalent to $P(Z^2 \geq z^2_{\alpha/2p}) = (\alpha/p)$ or $P\left(\chi^2(1) \geq z^2_{\alpha/2p}\right) = (\alpha/p)$. But $P\left(\chi^2(1) \geq \chi^2(1,\alpha/p)\right) = (\alpha/p)$. Thus the percentage points $z^2_{\alpha/2p}$ and $\chi^2(1,\alpha/p)$ are equivalent, and it is only necessary to compare $\chi^2(1,\alpha/p)$ with $\chi^2(p,\alpha)$ to determine which intervals are better. A computer was utilized to generate the percentage points $\chi^2(1,\alpha/p)$ and $\chi^2(p,\alpha)$ for selected values of α and p in order to determine whether any generalizations can be made regarding the superiority of one interval over the other. These are given in Table 3. Examination of Table 3 reveals that for $\alpha \leq .50$, actually for $\alpha \leq .55$, $\chi^2(1,\alpha/p) \leq \chi^2(p,\alpha)$ for all p, with equality holding for $p = 1$. Furthermore, the equality holds for all $0 < \alpha < 1$, not just $\alpha \leq .55$. From the above statement that

Table 3. Values of $\chi^2(1,\alpha/p)$ and $\chi^2(p,\alpha)$.

	α = .05		α = .10		α = .15	
p	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$
1	3.84	3.84	2.70	2.70	2.07	2.07
2	5.02	5.99	3.84	4.60	3.17	3.79
3	5.73	7.81	4.53	6.25	3.84	5.32
4	6.24	9.49	5.02	7.78	4.33	6.74
5	6.63	11.07	5.41	9.24	4.71	8.11
6	6.96	12.59	5.73	10.64	5.02	9.45
7	7.23	14.07	6.00	12.02	5.29	10.75
8	7.47	15.51	6.24	13.36	5.52	12.03
9	7.69	16.92	6.44	14.68	5.73	13.29
10	7.88	18.31	6.63	15.99	5.91	14.53
11	8.05	19.67	6.80	17.27	6.08	15.77
12	8.21	21.03	6.96	18.55	6.24	16.99
13	8.35	22.36	7.10	19.81	6.38	18.20
14	8.49	23.68	7.23	21.06	6.51	19.41
15	8.61	24.99	7.36	22.31	6.63	20.60
16	8.73	26.30	7.47	23.54	6.75	21.79
17	8.84	27.59	7.58	24.77	6.86	22.98
18	8.94	28.87	7.69	25.99	6.96	24.15
19	9.04	30.14	7.78	27.20	7.05	25.33
20	9.14	31.41	7.88	28.41	7.15	26.50
50	10.82	67.50	9.54	63.17	8.80	60.34

Table 3. Continued

p	α = .20		α = .25		α = .30	
	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$
1	1.64	1.64	1.32	1.32	1.07	1.07
2	2.70	3.22	2.35	2.77	2.07	2.41
3	3.36	4.64	3.00	4.10	2.70	3.66
4	3.84	5.99	3.47	5.38	3.17	4.88
5	4.22	7.29	3.84	6.62	3.54	6.06
6	4.53	8.56	4.15	7.84	3.84	7.23
7	4.79	9.80	4.41	9.04	4.10	8.38
8	5.02	11.03	4.64	10.22	4.33	9.52
9	5.23	12.24	4.84	11.39	4.53	10.66
10	5.41	13.44	5.02	12.55	4.71	11.78
11	5.58	14.63	5.19	13.70	4.87	12.90
12	5.73	15.81	5.34	14.84	5.02	14.01
13	5.87	16.98	5.48	15.98	5.16	15.12
14	6.00	18.15	5.61	17.12	5.29	16.22
15	6.12	19.31	5.73	18.24	5.41	17.32
16	6.24	20.46	5.84	19.37	5.52	18.42
17	6.34	21.61	5.95	20.49	5.63	19.51
18	6.44	22.76	6.05	21.60	5.73	20.60
19	6.54	23.90	6.14	22.72	5.82	21.69
20	6.63	25.04	6.24	23.83	5.91	22.77
50	8.28	58.16	7.88	56.33	7.55	54.72

Table 3. Continued

p	α = .35		α = .40		α = .45	
	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$
1	0.87	0.87	0.71	0.71	0.57	0.57
2	1.84	2.10	1.64	1.83	1.47	1.60
3	2.46	3.28	2.25	2.95	2.07	2.64
4	2.92	4.44	2.70	4.04	2.52	3.69
5	3.28	5.57	3.06	5.13	2.87	4.73
6	3.58	6.69	3.36	6.21	3.17	5.76
7	3.84	7.81	3.62	7.28	3.42	6.80
8	4.06	8.91	3.84	8.35	3.64	7.83
9	4.26	10.01	4.04	9.41	3.84	8.86
10	4.44	11.10	4.22	10.47	4.02	9.89
11	4.61	12.18	4.38	11.53	4.18	10.92
12	4.76	13.27	4.53	12.58	4.33	11.95
13	4.89	14.34	4.66	13.63	4.46	12.97
14	5.02	15.42	4.79	14.68	4.59	14.00
15	5.14	16.49	4.91	15.73	4.71	15.02
16	5.25	17.56	5.02	16.78	4.82	16.04
17	5.36	18.63	5.13	17.82	4.92	17.06
18	5.46	19.70	5.23	18.87	5.02	18.09
19	5.55	20.76	5.32	19.91	5.12	19.11
20	5.64	21.83	5.41	20.95	5.20	20.13
50	7.27	53.26	7.03	51.89	6.82	50.59

Table 3. Continued

p	α = .50		α = .55		α = .60	
	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$
1	0.45	0.45	0.36	0.36	0.27	0.27
2	1.32	1.39	1.191	1.196	1.07*	1.02
3	1.91	2.36	1.77	2.11	1.64	1.87
4	2.35	3.36	2.20	3.05	2.07	2.75
5	2.70	4.35	2.55	3.99	2.42	3.65
6	3.00	5.35	2.84	4.95	2.70	4.57
7	3.25	6.34	3.09	5.91	2.95	5.49
8	3.47	7.34	3.31	6.88	3.17	6.42
9	3.66	8.34	3.51	7.84	3.36	7.36
10	3.84	9.34	3.68	8.81	3.54	8.29
11	4.00	10.34	3.84	9.78	3.69	9.24
12	4.15	11.34	3.99	10.75	3.84	10.18
13	4.28	12.34	4.12	11.73	3.97	11.13
14	4.41	13.34	4.25	12.70	4.10	12.08
15	4.53	14.34	4.36	13.68	4.22	13.03
16	4.64	15.34	4.47	14.65	4.33	13.98
17	4.74	16.34	4.58	15.63	4.43	14.94
18	4.84	17.34	4.68	16.61	4.53	15.89
19	4.93	18.34	4.77	17.59	4.62	16.85
20	5.02	19.34	4.86	18.57	4.71	17.81
50	6.63	49.33	6.46	48.10	6.31	46.86

*The asterisked entry indicates that $\chi^2(1,\alpha/p)>\chi^2(p,\alpha)$.

Table 3. Continued

p	α = .65		α = .70		α = .75	
	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$
1	0.20	0.20	0.15	0.15	0.10	0.10
2	0.97*	0.86	0.87*	0.71	0.79*	0.57
3	1.52	1.64	1.42	1.42	1.32*	1.21
4	1.95	2.47	1.84	2.19	1.74	1.92
5	2.29	3.32	2.18	3.00	2.07	2.67
6	2.58	4.20	2.46	3.83	2.35	3.45
7	2.82	5.08	2.70	4.67	2.59	4.25
8	3.04	5.97	2.92	5.53	2.81	5.07
9	3.23	6.88	3.11	6.39	3.00	5.90
10	3.40	7.78	3.28	7.27	3.17	6.74
11	3.56	8.69	3.44	8.15	3.32	7.58
12	3.71	9.61	3.58	9.03	3.47	8.44
13	3.84	10.53	3.72	9.92	3.60	9.30
14	3.96	11.45	3.84	10.82	3.72	10.16
15	4.08	12.38	3.96	11.72	3.84	11.04
16	4.19	13.31	4.06	12.62	3.95	11.91
17	4.29	14.24	4.17	13.53	4.05	12.79
18	4.39	15.17	4.26	14.44	4.15	13.67
19	4.48	16.11	4.36	15.35	4.24	14.56
20	4.57	17.04	4.44	16.26	4.33	15.45
50	6.17	45.61	6.04	44.31	5.91	42.94

*The asterisked entries indicate that $\chi^2(1,\alpha/p) > \chi^2(p,\alpha)$.

Table 3. Continued

	α = .80		α = .85		α = .90		α = .95	
p	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$	$\chi^2(1,\alpha/p)$	$\chi^2(p,\alpha)$
1	0.06	0.06	0.03	0.03	0.01	0.01	0.004	0.004
2	0.71*	0.45	0.63*	0.32	0.57*	0.21	0.51*	0.10
3	1.23*	1.00	1.15*	0.80	1.07*	0.58	1.00*	0.35
4	1.64	1.65	1.55*	1.37	1.47*	1.06	1.39*	0.71
5	1.97	2.34	1.88	1.99	1.80*	1.61	1.72*	1.14
6	2.25	3.07	2.16	2.66	2.07	2.20	1.99*	1.63
7	2.49	3.82	2.40	3.36	2.31	2.83	2.22*	2.17
8	2.70	4.59	2.61	4.08	2.52	3.49	2.43	2.73
9	2.89	5.38	2.80	4.82	2.70	4.17	2.62	3.32
10	3.06	6.18	2.96	5.57	2.87	4.86	2.79	3.94
11	3.22	6.99	3.12	6.34	3.03	5.58	2.94	4.57
12	3.36	7.81	3.26	7.11	3.17	6.30	3.08	5.23
13	3.49	8.63	3.39	7.90	3.30	7.04	3.21	5.89
14	3.62	9.47	3.52	8.70	3.42	7.79	3.33	6.57
15	3.73	10.31	3.63	9.50	3.54	8.55	3.45	7.26
16	3.84	11.15	3.74	10.31	3.64	9.31	3.55	7.96
17	3.94	12.00	3.84	11.12	3.74	10.08	3.65	8.67
18	4.04	12.86	3.94	11.95	3.84	10.86	3.75	9.39
19	4.13	13.71	4.03	12.77	3.93	11.65	3.84	10.12
20	4.22	14.58	4.11	13.60	4.02	12.44	3.93	10.85
50	5.80	41.45	5.69	39.75	5.59	37.69	5.50	34.76

*The asterisked entries indicate that $\chi^2(1,\alpha/p)>\chi^2(p,\alpha)$.

$P\{\chi^2(1) \geq \chi^2(1,\alpha/p)\} = \alpha/p$, it immediately follows that $\chi^2(1,\alpha/p)$ equals $\chi^2(p,\alpha)$ for $p = 1$. If $\chi^2(1,\alpha/p) \leq \chi^2(p,\alpha)$ for $\alpha \leq .55$, then one must show that $P\{\chi^2(p) \geq \chi^2(1,\alpha/p)\} \geq \alpha$. This is the basis of the following conjecture.

Conjecture 4.1. For $\alpha \leq .55$ and $p \geq 1$, $z_{\alpha/2p} \leq \{\chi^2(p,\alpha)\}^{1/2}$.
In most practical applications, α is chosen to be much less than .55. Thus, if the conjecture is true, the Bonferroni intervals are preferable to the Scheffe intervals.

At this time, no proof is available.

Both types of intervals permit the decision maker to look at other confidence intervals than those on the p coordinate means. The Scheffe intervals easily offer this flexibility by varying the elements of the $\underset{\sim}{a}$ vector. Differences and other combinations of the mean vector can easily be analyzed in this fashion. For k linear combinations, the Bonferroni intervals can be found by introducing the appropriate $\underset{\sim}{a}$ in

$$\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\bar{\underset{\sim}{x}} \pm (\sqrt{n}\, z_{\alpha/2k})(\underset{\sim}{a}'\Sigma\underset{\sim}{a})^{1/2}.$$

This follows since $\bar{\underset{\sim}{X}} \sim N(\underset{\sim}{\mu},\Sigma/n)$ and $\underset{\sim}{a}'\bar{\underset{\sim}{X}} \sim N(\underset{\sim}{a}'\underset{\sim}{\mu},\underset{\sim}{a}'\Sigma\underset{\sim}{a}/n)$. In deciding which type of interval to use, the decision maker should compare $z_{\alpha/2k}$ with $\{\chi^2(p,\alpha)\}^{1/2}$ and determine which percentage point will provide for shorter intervals. The comparison of percentage points before construction of the intervals does not violate any statistical principles.

It should be pointed out that neither of these techniques make full use of the covariance structure but only select the diagonal elements. The extent to which confidence intervals may be shortened if the correlations are used is an avenue of future research.

Dunn (19) has obtained intervals of exact confidence level. The interval estimates are of the form $z_h \pm \sqrt{(p/n)}\, \sigma_h c_\alpha$, for $h = 1,\ldots,p$, where c_α is such that $P(Z \le c_\alpha) = \left(1+(1-\alpha)^{1/p}\right)/2$ and $z_h = \sum_{j=1}^{n} x_{hj} a_{jh}$; furthermore, $[x_{hj}]$ is the $(p\times n)$ data matrix and $[a_{jh}]$ is the $(n\times p)$ matrix formed by letting $A = BCD$, where the $(n\times n)$ orthogonal matrix B has its first column entries equal to $n^{-1/2}$, the $(n\times n)$ matrix C consists of a $(p\times p)$ orthogonal matrix in the upper left corner and zeros elsewhere, and the $(n\times p)$ matrix D consists of a $(p\times p)$ diagonal matrix with entries (p/n) on the main diagonal and zeros elsewhere and the last $(n-p)$ rows have all zero entries. However, as Dunn states, "The regions with exact confidence level are everywhere unnecessarily long." Dunn has also obtained the shortest possible intervals for the means provided $p = 2,3$, or $\rho_{ij} = b_i b_j$ for $i,j = 1,2,\ldots,p$, $i \neq j$ and with $0 < b_j < 1$. These intervals are of the form

$$\bar{x}_h \pm c_\alpha(\sigma_h/\sqrt{n}),$$

for $h = 1,2,\ldots,p$, where c_α is such that $P(Z \le c_\alpha) = \left(1+(1-\alpha)^{1/p}\right)/2$. However, an empirical investigation by Dunn for $p = 1(1)8$ reveals that c_α differs from the Bonferroni percentage points only in the second decimal place, if at all.

The first use of the χ^2 random variable in a control chart setting was by Hotelling (36,37) in the testing of bombsights. Let X_1 and X_2 be the range (measured in the direction of flight of the airplane) and deflection (measured perpendicular to the direction of flight) errors, respectively, for the dropping of a bomb at a target. Assume $\underset{\sim}{X} \sim N(\underset{\sim}{0}, \Sigma)$, where Σ is known. Let $\Lambda = [\lambda_{ij}]$ denote the inverse of the covariance matrix Σ. Then, for a single bomb,

$$\chi_B^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \lambda_{ij} X_{iB} X_{jB},$$

which is distributed as $\chi^2(2)$, is a measure of the quality of this bomb. Bombsights are tested in the following manner. From each lot of 20 sights, 2 sights are randomly chosen. Each of these sights is used on two flights with four bombs being dropped per flight. An overall measure of the quality of a flight(F), sight(S), or lot(L) is obtained by summing χ_B^2 over the number of bombs involved for that characteristic. This overall measure is denoted by χ_0^2. Thus,

$$\chi_{0F}^2 = \sum_{h=1}^{4} \chi_{B_h}^2, \quad \chi_{0S}^2 = \sum_{h=1}^{2} \chi_{0F_h}^2, \quad \text{and} \quad \chi_{0L}^2 = \sum_{h=1}^{2} \chi_{0S_h}^2,$$

where $\chi_{0F}^2 \sim \chi^2(8)$, $\chi_{0S}^2 \sim \chi^2(16)$, and $\chi_{0L}^2 \sim \chi^2(32)$. The overall measure can be partitioned into the mean point of impact (M) and the dispersion about the mean point of impact (D). That is, for n bombs,

$$\chi_M^2 = n \sum_{i=1}^{2} \sum_{j=1}^{2} \lambda_{ij}\bar{X}_i\bar{X}_j \quad \text{and} \quad \chi_D^2 = \sum_{h=1}^{n} \sum_{i=1}^{2} \sum_{j=1}^{2} \lambda_{ij}(X_{iB_h}-\bar{X}_i)(X_{jB_h}-\bar{X}_j).$$

Hotelling did not actually use χ^2 control charts since Σ was not known and his papers are principally devoted to this case. However, he does point out that control limits can be established for χ_0^2, χ_M^2, and χ_D^2 by using χ^2 percentage points when Σ is known.

Ghare and Torgersen (23) have also suggested the use of a χ^2 chart. However, their use is apparently incorrect since Σ was not known.

This chapter commenced by introducing a univariate χ^2 chart and emphasizing its relation to repeated tests of significance. At this point, these concepts were extended to the case of $p \geq 2$, and Hotelling's sparse treatment was elaborated upon. The concept of power for theoretical charts for the mean was also presented. Duncan (18) has presented the power for univariate charts. Furthermore, the use of simultaneous techniques in conjunction with control charts was also presented. Finally, Table 3 and Conjecture 4.1 suggest that the Bonferroni intervals are superior to the Scheffe intervals for $\alpha < .50$. Although the procedure for testing H_0: $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs. H_1: $\underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ with Σ known has been extensively treated previous to this research, this procedure has apparently not been applied to a control chart environment, except as noted. This includes the concept of the power and the simultaneous techniques.

CHAPTER V

EMPIRICAL CONTROL CHARTS FOR THE MEAN

In Chapter IV, a process was assumed to be in a state of statistical control at a selected standard value of μ_0, and a χ^2 control chart was established to assist the decision maker in determining whether the process will remain stable. The state of control was completely defined since Σ was also specified. In contrast to these *theoretical* control charts are *empirical* control charts which are used for analyzing the lack of control of past operations and to assist in establishing theoretical control charts. In this situation, a satisfactory state of statistical control may not have yet been established, and the parameters of the distribution from which the sample is drawn are unknown. Thus, the control limits cannot be determined *a priori*.

1. Empirical Control Charts for Rational Subgroups

A common procedure for analyzing past operations is based on the concept of rational subgrouping, which assumes each subgroup of observations was taken under identical conditions. To establish empirical or estimated control charts, k homogeneous samples of size n are selected, from which certain statistics are computed and used as estimates in the control limits of the corresponding theoretical control charts. These estimated control charts are then used to analyze the subgroups of past observations for lack of control. These charts are also used for controlling the quality of future observations. If the past data indicate lack of control, certain causes may have been present and contributed to the process being unstable. If these principal assignable causes are detected and removed and new control limits are established, then the remaining past measurements should behave as coming from a stable distribution. In order to detect assignable causes, auxiliary observations regarding physical factors usually need to have been recorded along with the numerical data.

One Quality Characteristic

For p = 1, let μ_0 and σ^2 denote the unknown mean and variance, respectively, of a stable process. Assume that k rational subgroups of n observations each have been collected, where statistical control existed within each subgroup but not necessarily between the k subgroups. If the within subgroup population variance is the same for all subgroups, i.e., $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$, the process will be out of control only if the subgroup population means are different. Refer to Table 4.

If the process is in a state of control, $\bar{X}_h \sim N(\mu_0, \sigma^2/n)$ for $h = 1,2,\ldots,k$. The 3-sigma *theoretical* control chart values were given by

$$UCL = \mu_0 + 3(\sigma/\sqrt{n})$$

$$CL = \mu_0$$

$$LCL = \mu_0 - 3(\sigma/\sqrt{n})$$

Table 4. Data for p = 1

Sample Number	Individual Values	Subgroup Statistics	Population Parameters
1	$x_{11}, x_{12}, \ldots, x_{1n}$	$\bar{x}_1, s_1^2, s_1$	μ_1, σ^2
2	$x_{21}, x_{22}, \ldots, x_{2n}$	$\bar{x}_2, s_2^2, s_2$	μ_2, σ^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$x_{k1}, x_{k2}, \ldots, x_{kn}$	$\bar{x}_k, s_k^2, s_k$	μ_k, σ^2

Since μ_0 and σ^2 are unknown, it would seem reasonable to replace them by their estimates. Unbiased estimates of μ_0 and σ^2, denoted $\bar{\bar{x}}$ and s_p^2, respectively, are given by $\bar{\bar{x}} = (1/k) \sum_{h=1}^{k} \bar{x}_h$ and $s_p^2 = (1/k) \sum_{h=1}^{k} s_h^2$, where $\bar{x}_h = (1/n) \sum_{j=1}^{n} x_{hj}$ and $s_h^2 = (n-1)^{-1} \sum_{j=1}^{n} (x_{hj}-\bar{x}_h)^2$. Note that $(n-1)S_h^2/\sigma^2$ is distributed $\chi^2(n-1)$ and, thus, $(1/\sigma^2) \sum_{h=1}^{k} (n-1)S_h^2$ is distributed $\chi^2[k(n-1)]$. Thus, using the reasonable criterion of replacing the population parameters by their unbiased estimates, the *empirical* 3-sigma control chart values for $\bar{x}_h$ are given by

$$\mathrm{UCL} = \bar{\bar{x}} + A_1 s_p^*$$

$$\mathrm{CL} = \bar{\bar{x}}$$

$$\mathrm{LCL} = \bar{\bar{x}} - A_1 s_p^*,$$

where A_1 is chosen so that $E(A_1 S_p^*) = (3\sigma/\sqrt{n})$ and $S_p^* = (1/k)\sum_{h=1}^{k} S_h$. Note that S_p^* was used in place of S_p and S_p^* is not defined to be $S_p = \sqrt{S_p^2}$. Furthermore, $E(S_h) \neq \sigma$ even though $E(S_h^2) = \sigma^2$, for h = 1,2,...,k. In fact, $E(S_h) = \sigma c_2'$, where $c_2' = \{2/(n-1)\}^{1/2}\Gamma(n/2)/\Gamma\{(n-1)/2\}$. The traditional quality control literature gives different values for the correction factors since V_h is used for S_h, where $V_h^2 = n^{-1}\sum_{j=1}^{n}(X_{hj}-\bar{X}_h)^2$. In this case, the correction factor is denoted c_2 and is given by $c_2 = (2/n)^{1/2}\Gamma(n/2)/\Gamma\{(n-1)/2\}$. Note, $c_2 = c_2'(1-n^{-1})^{1/2}$, and $A_1 = 3/(c_2'\sqrt{n})$. Johnson and Leone (46) give values of both c_2' and c_2 for n = 2(1)25. Tables of A_1 are given in Guttman and Wilks (27) for n = 2(1)10. Both of these texts are apparently in error regarding the following minor point.

Since, by definition, $S_p = \sqrt{S_p^2}$, it is obvious that $S_p \neq S_p^*$. In fact, $S_p > S_p^*$, and this is easily shown for k = 2. Since $(S_1-S_2)^2 = S_1^2 + S_2^2 - 2S_1S_2 > 0$, it follows that $S_1^2 + S_2^2 > 2S_1S_2$ and $2(S_1^2+S_2^2) > 2S_1S_2 + S_1^2 + S_2^2 = (S_1+S_2)^2$. Thus, $S_p > S_p^*$. Furthermore, $E(S_p) = \sigma c_2''$, where

$$c_2'' = \{2/(kn-k)\}^{1/2}\Gamma\{(kn-k+1)/2\}/\Gamma\{(kn-k)/2\}.$$

In calculating the control limits for either examples or exercises, both texts seem to use S_p in conjunction with c_2' as the correction factor, when apparently S_p^* should be used with c_2'. Chapter VII will present additional material regarding c_2 and c_2'.

The control of the process is tested by plotting the k sample means on the estimated 3-sigma control chart. If one or more of the subgroup sample means plot outside the estimated control limits, it is hopeful that assignable causes of variation can be detected and removed. In this manner, a state of control is eventually approached.

Determining the control of the process could also be viewed as a hypothesis testing problem. One would set up the null hypothesis H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ against all possible alternatives. Specifically, Ω and ω are those subsets of Euclidean k + 1 space such that

$$\Omega = \{(\mu_1,\mu_2,\ldots,\mu_k,\sigma^2): -\infty<\mu_h<\infty,\ \sigma^2>0,\ h = 1,2,\ldots,k\}$$

and

$$\omega = \{(\mu_1,\mu_2,\ldots,\mu_k,\sigma^2): -\infty<\mu_1 = \mu_2 = \ldots = \mu_k<\infty,\ \sigma^2>0\}.$$

Let $L(\omega)$ and $L(\Omega)$ denote the likelihood functions for the parameter spaces ω and Ω, respectively. Also let $L(\hat{\omega})$ and $L(\hat{\Omega})$ denote the supremum of these functions. Then, if μ denotes the common but unknown value of $\mu_1,\mu_2,\ldots,\mu_k$,

$$L(\omega) = (2\pi\sigma^2)^{-nk/2}\exp\left(-(1/2\sigma^2)\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\mu)^2\right)$$

and

$$L(\Omega) = (2\pi\sigma^2)^{-nk/2}\exp\left(-(1/2\sigma^2)\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\mu_i)^2\right).$$

By obtaining $\partial \ln L(\omega)/\partial\mu$ and $\partial \ln L(\omega)/\partial\sigma^2$, equating the results to zero, and substituting the results in $L(\omega)$, one obtains

$$L(\hat{\omega}) = \left[\frac{nk}{2\pi \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij}-\bar{\bar{x}})^2}\right]^{nk/2} e^{-nk/2}.$$

Similarly, one obtains

$$L(\hat{\Omega}) = \left[\frac{nk}{2\pi \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij}-\bar{x}_i)^2}\right]^{nk/2} e^{-nk/2}.$$

Thus,

$$\lambda(\underset{\sim}{x}') = \left[\frac{\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij}-\bar{x}_i)^2}{\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij}-\bar{\bar{x}})^2}\right]^{nk/2},$$

where $\underset{\sim}{x}' = [x_{11},\ldots,x_{1n},x_{21},\ldots,x_{2n},\ldots,x_{k1},\ldots,x_{kn}]$. However, $\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij}-\bar{\bar{x}})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij}-\bar{x}_i)^2 + n \sum_{i=1}^{k} (\bar{x}_i-\bar{\bar{x}})^2$, and the form of the critical region is given by

$$w = \left\{\underset{\sim}{x}' : \frac{n \sum_{i=1}^{k} (\bar{x}_i-\bar{\bar{x}})^2/(k-1)}{\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij}-\bar{x}_i)^2/k(n-1)} > c\right\}, \qquad (17)$$

where c is usually chosen so that the probability of type I error is α.

Let $F(m_1,m_2)$ denote Snedecor's F random variable with m_1 and m_2 degrees of freedom and let $F(m_1,m_2,\alpha)$ denote the upper α-percentile point of this random variable, where $F(m_1,m_2) = \left(m_2\chi^2(m_1)\right)/\left(m_1\chi^2(m_2)\right)$ and $\chi^2(m_1)$ and $\chi^2(m_2)$ are independent. Since $n \sum_{i=1}^{k} (\bar{X}_i-\bar{\bar{X}})^2/\sigma^2 \sim \chi^2(k-1)$ and $\sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij}-\bar{X}_i)^2/\sigma^2 \sim \chi^2\left(k(n-1)\right)$, $c = F\left(k-1,k(n-1),\alpha\right)$. Thus, the decision maker is simultaneously testing $\mu_1 = \mu_2 = \ldots = \mu_k$ versus not all μ_h's are equal with significance level α, which is the familiar one-way analysis of variance problem. When $k = 2$, this reduces to the problem of testing H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$, which is based on the t statistic with $(2n-2)$ d.f.

In testing H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ or H_0: $\mu_i = \mu_j$ for $i \neq j$, the conclusion may be to reject this hypothesis. The decision maker is now interested in those pairs of means which are not equal. One way of simultaneously analyzing these $\binom{k}{2}$ pairs of means is by the use of Scheffe's F technique. This technique and others are thoroughly presented by Miller (53). Because the necessary background has already been presented, Scheffe's technique is stated in the following remark.

Remark 5.1. Consider the k independent random variables $X_1,\ldots,X_k$, where $X_h \sim N(\mu_h,\sigma^2)$, with both μ_h and σ^2 being unknown. Let a random sample of size n be taken of each random variable, where $X_{h1},X_{h2},\ldots,X_{hn}$ denote the elements of the random sample of X_h, $h = 1,2,\ldots,k$. Let $S_p^2 = \left(k(n-1)\right)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij}-\bar{X}_i)^2$. Then, letting $F_\alpha = F\left(k,k(n-1),\alpha\right)$,

$$P[\underset{\sim}{a}'\bar{\underset{\sim}{X}}-\|\underset{\sim}{a}\|\sqrt{S_p^2(k/n)F_\alpha}\le\underset{\sim}{a}'\underset{\sim}{\mu}\le\underset{\sim}{a}'\bar{\underset{\sim}{X}}+\|\underset{\sim}{a}\|\sqrt{S_p^2(k/n)F_\alpha}\text{ for all nonnull }\underset{\sim}{a}] = 1 - \alpha .$$

Since the decision maker would only be interested in a finite number of the $\underset{\sim}{a}$, namely, where the only two nonzero coordinates of $\underset{\sim}{a}$ are +1 and -1, the confidence coefficient is at least $1 - \alpha$. The decision maker could also obtain confidence intervals on μ_h by letting $\underset{\sim}{a} = \underset{\sim}{e}_h$.

Thus, in establishing control of the mean, the decision maker could analyze the data for lack of control by the reasonable procedure of control charts or the hypothesis testing anova technique in conjunction with the simultaneous procedure of Scheffe. Which one to use is a matter of preference. C. C. Craig (14) has summarized the various aspects of each, and Duncan (18) reiterates some of these in his book. Craig states that ". . . the two methods are nearly enough equivalent that both will disclose any clear lack of control among averages and when one says control is good the other will too."

After the initial k samples have been selected and their control or lack of it has been determined, the customary practice is to use the empirical chart in controlling future samples until such a time as the decision maker feels it is wise to convert to theoretical control limits. Of course, this is not an exact statistical procedure.

Multiple Quality Characteristics

A natural generalization of empirical control charts for rational subgroups is when $p > 1$. However, a necessary prerequisite is the statement and derivation of Hotelling's T^2 distribution under general conditions. The general distribution was first derived by Hsu (38). The following statement and derivation are based on the presentations by Bowker (7) and Anderson (3). A representation is deemed necessary due to its frequent misuse. First consider the noncentral F distribution.

Definition 5.1. If a random variable is distributed as $\chi'^2(m_1,\lambda)$ and if another random variable is distributed as $\chi^2(m_2)$, where $\chi'^2(m_1,\lambda)$ and $\chi^2(m_2)$ are independent, then the random variable

$$\frac{\chi'^2(m_1,\lambda)/m_1}{\chi^2(m_2)/m_2} = \frac{m_2\chi'^2(m_1,\lambda)}{m_1\chi^2} \tag{18}$$

has the noncentral F distribution with m_1,m_2 degrees of freedom and noncentrality parameter λ, denoted by $F'(m_1,m_2,\lambda)$.

$F'(m_1,m_2,\lambda,\alpha)$ will denote the upper α percentile point of $F'(m_1,m_2,\lambda)$. When $\lambda = 0$, $F'(m_1,m_2,\lambda=0) = F(m_1,m_2)$.

Since $f_{\chi'^2(m_1,\lambda)}$ was derived in Chapter IV, $f_{F'(m_1,m_2,\lambda)}$ can easily be derived by making the appropriate transformations. See Graybill (25). The following theorem describes Hotelling's T^2 distribution.

Theorem 5.1. Let $T^2 = \underset{\sim}{U}'S^{-1}\underset{\sim}{U}$, where $\underset{\sim}{U}\sim N(\underset{\sim}{\mu},\Sigma)$ and νS is independently distributed as $\sum_{h=1}^{\nu} \underset{\sim}{Y}_h\underset{\sim}{Y}_h'$ with the $\underset{\sim}{Y}_h$ independent and $\underset{\sim}{Y}_h\sim N(\underset{\sim}{0},\Sigma)$, $h = 1,2,\ldots,\nu$. Then $(T^2/\nu)\big((\nu-p+1)/p\big)$ is distributed as $F'(p,\nu-p+1,\lambda=\underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu})$.

Proof. There exists a nonsingular $(p\times p)$ matrix R such that $\Sigma = RR'$. Define $\underset{\sim}{U}^* = R^{-1}\underset{\sim}{U}$, $S^* = R^{-1}S(R^{-1})'$, $\underset{\sim}{\mu}^* = R^{-1}\underset{\sim}{\mu}$. Then $T^2 = \underset{\sim}{U}'S^{-1}\underset{\sim}{U} =$

$\underset{\sim}{U}'(R^{-1})'R'S^{-1}RR^{-1}\underset{\sim}{U} = \underset{\sim}{U}^{*'}S^{*-1}\underset{\sim}{U}^{*}$, where $\underset{\sim}{U}^{*} \sim N(\underset{\sim}{\mu}^{*}, I)$ and νS^{*} is independently distributed as $\sum_{h=1}^{\nu} \underset{\sim}{Y}_h^{*}\underset{\sim}{Y}_h^{*'}$ with the $\underset{\sim}{Y}_h^{*} = R^{-1}\underset{\sim}{Y}_h$ independent and $\underset{\sim}{Y}_h^{*} \sim N(\underset{\sim}{0}, I)$. Let H be a (p×p) orthogonal matrix whose first row is given by $h_{1j} = \mu_j^{*}/\sqrt{\underset{\sim}{\mu}^{*'}\underset{\sim}{\mu}^{*}}$, $j = 1,2,\ldots,p$. Define $\underset{\sim}{V} = H\underset{\sim}{U}^{*}$, $A = H(\nu S^{*})H'$. Then $T^2 = \underset{\sim}{U}^{*'}S^{*-1}\underset{\sim}{U}^{*} = \underset{\sim}{V}'HH'(\nu A^{-1})HH'\underset{\sim}{V} = \nu\underset{\sim}{V}'A^{-1}\underset{\sim}{V}$, where $\underset{\sim}{V} \sim N(H\underset{\sim}{\mu}^{*}, I)$ and A is independently distributed as $\sum_{h=1}^{\nu} \underset{\sim}{Y}_h^{*}\underset{\sim}{Y}_h^{*'}$. Note that $H\underset{\sim}{\mu}^{*} = \left(\sqrt{\underset{\sim}{\mu}^{*'}\underset{\sim}{\mu}^{*}}, 0, \ldots, 0\right)'$. For any set of values of $\underset{\sim}{V}$ (except the null vector which has probability zero) define a (p×p) random orthogonal matrix B with first row given by $b_{1j} = V_i/\sqrt{\underset{\sim}{V}'\underset{\sim}{V}}$, $j = 1,2,\ldots,p$. Then $T^2 = \nu\underset{\sim}{V}'A^{-1}\underset{\sim}{V} = \nu\underset{\sim}{V}'B'BA^{-1}B'B\underset{\sim}{V} = \nu(B\underset{\sim}{V})'(BAB')^{-1}(B\underset{\sim}{V})$. Let $A^{*} = BAB'$ and note that $B\underset{\sim}{V} = (\sqrt{\underset{\sim}{V}'\underset{\sim}{V}}, 0, \ldots, 0)'$. Then $T^2 = \nu(\sqrt{\underset{\sim}{V}'\underset{\sim}{V}}, 0, \ldots, 0)(A^{*})^{-1}(\sqrt{\underset{\sim}{V}'\underset{\sim}{V}}, 0, \ldots, 0)' = \nu\sqrt{\underset{\sim}{V}'\underset{\sim}{V}}(a^{*11})\sqrt{\underset{\sim}{V}'\underset{\sim}{V}} = \nu(\underset{\sim}{V}'\underset{\sim}{V})(a^{*11})$, where a^{*11} denotes the element in the first row and first column of $(A^{*})^{-1}$, and $a^{*11} = 1/(a^{*11})^{-1}$. Thus, $(T^2/\nu) = (\underset{\sim}{V}'\underset{\sim}{V})/(a^{*11})^{-1}$. By Lemma 3.6, A^{*} has a Wishart distribution and its elements are distributed independently of the elements of B and consequently of $\underset{\sim}{V}'\underset{\sim}{V}$. By Lemma 3.7, $(a^{*11})^{-1}$ has density function $w\left((a^{*11})^{-1} \mid 1, 1, \nu-p+1\right)$, which is a chi-square density with $(\nu-p+1)$ degrees of freedom. In view of Definition 4.1, $\underset{\sim}{V}'\underset{\sim}{V}$ is distributed as $\chi'^2(p,\lambda)$ with $\lambda = (H\underset{\sim}{\mu}^{*})'(H\underset{\sim}{\mu}^{*}) = \underset{\sim}{\mu}^{*'}\underset{\sim}{\mu}^{*} = (R^{-1}\underset{\sim}{\mu})'(R^{-1}\underset{\sim}{\mu}) = \underset{\sim}{\mu}'(RR')^{-1}\underset{\sim}{\mu} = \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu}$. Then (T^2/ν) is distributed as the ratio of the independent random variables $\chi'^2(p,\lambda)$ and $\chi^2(\nu-p+1)$. Thus, $(T^2/\nu)\left\{(\nu-p+1)/p\right\} = F'(p, \nu-p+1, \lambda=\underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu})$. ‖

The multivariate analogue of empirical control charts for more than one quality characteristic when there are rational subgroups may now be considered. Assume that k rational subgroups have been defined, where statistical control existed within each subgroup but not necessarily between the k subgroups. If the covariance matrix Σ is common to all k groups, i.e., $\Sigma^{(1)} = \Sigma^{(2)} = \ldots = \Sigma^{(k)} = \Sigma$, the process will be out of control only if the subgroup population means are different. Refer to Table 5.

Table 5. Data for $p > 1$

Sample Number	Data Matrices	Subgroup Statistics	Population Parameters
1	$X^{(1)}$	$\bar{\underset{\sim}{x}}^{(1)}, S^{(1)}$	$\underset{\sim}{\mu}^{(1)}, \Sigma$
2	$X^{(2)}$	$\bar{\underset{\sim}{x}}^{(2)}, S^{(2)}$	$\underset{\sim}{\mu}^{(2)}, \Sigma$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$X^{(k)}$	$\bar{\underset{\sim}{x}}^{(k)}, S^{(k)}$	$\underset{\sim}{\mu}^{(k)}, \Sigma$

In Table 5, consider group h. Then for $h = 1,2,\ldots,k$,

$$X^{(h)} = \begin{bmatrix} x_{11}^{(h)} & x_{12}^{(h)} & \cdots & x_{1n}^{(h)} \\ \vdots & \vdots & & \vdots \\ x_{p1}^{(h)} & x_{p2}^{(h)} & \cdots & x_{pn}^{(h)} \end{bmatrix} = \left[\underset{\sim}{x}_1^{(h)}, \underset{\sim}{x}_2^{(h)}, \ldots, \underset{\sim}{x}_n^{(h)}\right],$$

$$\bar{\underset{\sim}{x}}^{(h)} = \begin{bmatrix} (1/n) \sum_{j=1}^{n} x_{1j}^{(h)} \\ \vdots \\ (1/n) \sum_{j=1}^{n} x_{pj}^{(h)} \end{bmatrix} = \begin{bmatrix} \bar{x}_1^{(h)} \\ \vdots \\ \bar{x}_p^{(h)} \end{bmatrix},$$

$$A^{(h)} = \begin{bmatrix} \sum_{j=1}^{n} (x_{1j}^{(h)}-\bar{x}_1^{(h)})^2 & \cdots & \sum_{j=1}^{n} (x_{1j}^{(h)}-\bar{x}_1^{(h)})(x_{pj}^{(h)}-\bar{x}_p^{(h)}) \\ \vdots & & \vdots \\ \sum_{j=1}^{n} (x_{pj}^{(h)}-\bar{x}_p^{(h)})(x_{1j}^{(h)}-\bar{x}_1^{(h)}) & \cdots & \sum_{j=1}^{n} (x_{pj}^{(h)}-\bar{x}_p^{(h)})^2 \end{bmatrix},$$

and

$$S^{(h)} = (n-1)^{-1} A^{(h)}.$$

If the process is in a state of control, $\bar{\underset{\sim}{X}}^{(h)} \sim N\left(\underset{\sim}{\mu}_0, (1/n)\Sigma\right)$, for $h = 1,2,\ldots,k$, where $\underset{\sim}{\mu}_0$ denotes the common population mean. For the case where Σ was known, the theoretical control limit was given by $\chi^2(p,\alpha)$ and the process was judged out of control if $n(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)$ exceeded $\chi^2(p,\alpha)$. Now both $\underset{\sim}{\mu}_0$ and Σ are unknown. It would seem reasonable to replace them by their unbiased estimates and make any other necessary adjustments. Unbiased estimates of $\underset{\sim}{\mu}_0$ and Σ, denoted $\bar{\bar{\underset{\sim}{x}}}$ and S_p, respectively, are given by $\bar{\bar{\underset{\sim}{x}}} = (1/k) \sum_{h=1}^{k} \bar{\underset{\sim}{x}}^{(h)}$ and $S_p = \left(k(n-1)\right)^{-1} \sum_{h=1}^{k} A^{(h)} = (1/k) \sum_{h=1}^{k} S^{(h)}$. Then the distribution of $c(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})' S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})$ needs to be determined, where c is some fixed constant.

For $h = 1,\ldots,k$, $\bar{\underset{\sim}{X}}^{(h)} \sim N\left(\underset{\sim}{\mu}_0, (1/n)\Sigma\right)$, and the $\bar{\underset{\sim}{X}}^{(h)}$ are independent; in addition, $\bar{\bar{\underset{\sim}{X}}} \sim N\left(\underset{\sim}{\mu}_0, (kn)^{-1}\Sigma\right)$. Also note that independence exists among the n vectors within a group and among the n vectors of all k subgroups. Let $\underset{\sim}{Y} = \bar{\underset{\sim}{X}}^{(h)} - \bar{\bar{\underset{\sim}{X}}}$. Then $\underset{\sim}{Y} = (1/n) \sum_{j=1}^{n} \underset{\sim}{X}_j^{(h)} - (1/kn) \sum_{i=1}^{k} \sum_{j=1}^{n} \underset{\sim}{X}_j^{(i)} = \left((k-1)/(kn)\right) \sum_{j=1}^{n} \underset{\sim}{X}_j^{(h)} - (1/kn) \sum_{i \neq h} \sum_{j=1}^{n} \underset{\sim}{X}_j^{(i)}$. Note that $E\underset{\sim}{Y} = \left((k-1)/(kn)\right)(n\underset{\sim}{\mu}_0) - (1/kn)(k-1)(n\underset{\sim}{\mu}_0) = \underset{\sim}{0}$. In expanded form, $\underset{\sim}{Y} = \left((k-1)/(kn)\right)\underset{\sim}{X}_1^{(h)} + \ldots + \left((k-1)/(kn)\right)\underset{\sim}{X}_n^{(h)} - (1/kn)\underset{\sim}{X}_1^{(1)} - \ldots - (1/kn)\underset{\sim}{X}_n^{(1)} - \ldots - (1/kn)\underset{\sim}{X}_1^{(h-1)} - \ldots - (1/kn)\underset{\sim}{X}_n^{(h-1)} - (1/kn)\underset{\sim}{X}_1^{(h+1)} - \ldots - (1/kn)\underset{\sim}{X}_n^{(h+1)} - \ldots - (1/kn)\underset{\sim}{X}_1^{(k)} - \ldots - (1/kn)\underset{\sim}{X}_n^{(k)}$. Consider any $(p\times 1)$ vector of constants $\underset{\sim}{a}$. Then $\underset{\sim}{a}'\underset{\sim}{Y} = \left((k-1)/(kn)\right)\underset{\sim}{a}'\underset{\sim}{X}_1^{(h)} + \ldots + \left((k-1)/(kn)\right)\underset{\sim}{a}'\underset{\sim}{X}_n^{(h)} - (1/kn)\underset{\sim}{a}'\underset{\sim}{X}_1^{(1)} - \ldots - (1/kn)\underset{\sim}{a}'\underset{\sim}{X}_n^{(1)} - \ldots - (1/kn)\underset{\sim}{a}'\underset{\sim}{X}_1^{(k)} - (1/kn)\underset{\sim}{a}'\underset{\sim}{X}_n^{(k)}$, and $\underset{\sim}{a}'\underset{\sim}{Y}$ is univariate normally distributed since it is a linear combination of kn independent, univariate normal random variables. Thus, $V(\underset{\sim}{a}'\underset{\sim}{Y}) = \left((k-1)/(kn)\right)^2 V(\underset{\sim}{a}'\underset{\sim}{X}_1^{(h)}) + \ldots + \left((k-1)/(kn)\right)^2 V(\underset{\sim}{a}'\underset{\sim}{X}_n^{(h)}) + (1/kn)^2 V(\underset{\sim}{a}'\underset{\sim}{X}_1^{(1)}) + \ldots + (1/kn)^2 V(\underset{\sim}{a}'\underset{\sim}{X}_n^{(1)}) + \ldots + (1/kn)^2 V(\underset{\sim}{a}'\underset{\sim}{X}_1^{(k)}) + (1/kn)^2 V(\underset{\sim}{a}'\underset{\sim}{X}_n^{(k)}) = \left((k-1)^2/(kn)^2\right)(n)(\underset{\sim}{a}'\Sigma\underset{\sim}{a}) + (1/kn)^2(k-1)(n)(\underset{\sim}{a}'\Sigma\underset{\sim}{a}) = \left((k-1)/(kn)\right)(\underset{\sim}{a}'\Sigma\underset{\sim}{a}) = \underset{\sim}{a}'\left((k-1)\Sigma/(kn)\right)\underset{\sim}{a}$, which indicates that $(k-1)\Sigma/(kn)$ is the covariance matrix of $\underset{\sim}{Y}$. Thus, $\underset{\sim}{Y} \sim N\left(\underset{\sim}{0}, (k-1)\Sigma/(kn)\right)$. Let $\underset{\sim}{U} = \sqrt{(kn)/(k-1)}\,\underset{\sim}{Y}$. Then $\underset{\sim}{U} \sim N(\underset{\sim}{0}, \Sigma)$. Now $S_p = \left(k(n-1)\right)^{-1} \sum_{h=1}^{k} A^{(h)} = \left(k(n-1)\right)^{-1}\left(\sum_{j=1}^{n-1} \underset{\sim}{Y}_j\underset{\sim}{Y}_j' + \ldots + \sum_{j=1}^{n-1} \underset{\sim}{Y}_j\underset{\sim}{Y}_j'\right) = \left(k(n-1)\right)^{-1}\left(\sum_{j=1}^{k(n-1)} \underset{\sim}{Y}_j\underset{\sim}{Y}_j'\right)$, where the $\underset{\sim}{Y}_j$ are independent and $\underset{\sim}{Y}_j \sim N(\underset{\sim}{0}, \Sigma)$, $j = 1,2,\ldots,k(n-1)$. Thus, $k(n-1)S_p$ is distributed as $\sum_{j=1}^{k(n-1)} \underset{\sim}{Y}_j\underset{\sim}{Y}_j'$. Also note that $(\bar{\underset{\sim}{X}}^{(h)} - \bar{\bar{\underset{\sim}{X}}})$ and S_p are independent. Then, by Theorem 5.1, for $h = 1,2,\ldots,k$, $\left((kn)/(k-1)\right)(\bar{\underset{\sim}{X}}^{(h)}-$

$\bar{\bar{X}})'S_p^{-1}(\bar{X}^{(h)}-\bar{\bar{X}})$ is distributed as $[k(n-1)p/(kn-k-p+1)]F(p,kn-k-p+1)$. Note that the noncentrality parameter is 0.

An empirical upper control limit is given by the percentage point $[k(n-1)p/(kn-k-p+1)]F(p,kn-k-p+1,\alpha)$, and there is only an upper limit since the quantities $[(kn)/(k-1)](\bar{X}^{(h)}-\bar{\bar{X}})'S_p^{-1}(\bar{X}^{(h)}-\bar{\bar{X}})$ are a generalized measure of distance. The control of the past process is tested by plotting the k quantities, $[(kn)/(k-1)](\bar{x}^{(h)}-\bar{\bar{x}})'S_p^{-1}(\bar{x}^{(h)}-\bar{\bar{x}})$, against the control limit on a chart similar to that illustrated in Figure 4. If one or more of the k quantities plots above the control limit, the process is judged out of control and assignable causes of variation are sought. If an assignable cause is found, then the data matrices for the out of control quantities are eliminated from consideration and a new upper control limit is found by an appropriate reduction in k. As in the univariate case, this control chart method is only an ad hoc method. The exact formulation is a hypothesis testing problem which is treated after consideration of the out of control quantities.

It may have been more natural to use $n(\bar{X}^{(h)}-\bar{\bar{X}})'S_p^{-1}(\bar{X}^{(h)}-\bar{\bar{X}})$, for $h = 1,2,\ldots,k$, as the test statistic since this more closely parallels $n(\bar{X}-\mu_0)'\Sigma^{-1}(\bar{X}-\mu_0)$, the theoretical control chart test statistic. If $n(\bar{X}^{(h)}-\bar{\bar{X}})'S_p^{-1}(\bar{X}^{(h)}-\bar{\bar{X}})$ is used, the empirical upper control limit is given by $c(k,n,p)F(p,kn-k-p+1,\alpha)$, where $c(k,n,p) = (k^2np-k^2p-knp+kp)/(k^2n-k^2-kp+k)$. It can be shown that $F(\nu_1,\nu_2)$ approaches a $\chi^2(\nu_1)/\nu_1$ as ν_2 approaches infinity. Furthermore, $c(k,n,p)$ approaches p as k approaches infinity. Thus, the upper control limit becomes $\chi^2(p,\alpha)$, the theoretical control chart limit, as the number of subgroups becomes infinitely large. This suggests that the empirical control chart can also be viewed as repeated tests of significance as the number of subgroups increases. It should be pointed out that, as $k \to \infty$, $c(k,n,p)F(p,kn-k-p+1,\alpha)$ *increases* to $\chi^2(p,\alpha)$. Even though $F(p,kn-k-p+1,\alpha)$ decreases to $\chi^2(p,\alpha)/p$, $c(k,n,p)$ increases to p, as $k \to \infty$. This is pointed out in Table 6 for $\alpha = .01$, $p = 2$, $n = 10$, and $k = 2(2)20,\infty$.

Table 6. Comparison of $c(k,n,p)F(p,kn-k-p+1,\alpha)$ with $\chi^2(p,\alpha)$ for $\alpha = .01$, $p = 2$, $n = 10$, and Selected k

k	c(k,10,2)	F(2,9k-1,.01)	c(k,10,2)F(9k-1,.01)	$\chi^2(2,.01)$
2	1.059	6.11	6.47	9.21
4	1.543	5.27	8.13	9.21
6	1.698	5.03	8.54	9.21
8	1.775	4.92	8.73	9.21
10	1.820	4.85	8.83	9.21
12	1.851	4.81	8.90	9.21
14	1.872	4.78	8.95	9.21
16	1.888	4.76	8.99	9.21
18	1.901	4.74	9.01	9.21
20	1.911	4.73	9.03	9.21
∞	2.000	4.605	9.21	9.21

However, the overall test statistic, which is $(T^2/\nu)(\nu-p+1/p)$ with $\nu = k(n-1)$, is distributed as an $F(p,\nu-p+1)$, and its percentage point is decreasing as k increases. Also note that the upper control limit becomes $(1-k^{-1})\chi^2(p,\alpha)$ as the number of observation vectors, n, in each subgroup approaches infinity. When there is only one quality characteristic, the control limit becomes $(1-k^{-1})F(1,kn-k,\alpha)$, where $F(1,kn-k)$ is equal to the square of Student's t with (kn-k) degrees of freedom.

This result can be related to the univariate empirical control chart in the following manner. The upper and lower empirical control limits were given by $\bar{\bar{x}} \pm A_1 s_p^*$, where $E(A_1 S_p^*) = (3\sigma)/\sqrt{n}$, $A_1 = 3/(c_2'\sqrt{n})$, and $c_2' = [2/(n-1)]^{1/2}\Gamma(n/2)/\Gamma[(n-1)/2]$. As previously stated, the control limits could have also been expressed using $S_p = \sqrt{S_p^2}$. In this case, the control limits would be given by $\bar{\bar{x}} \pm A_1' s_p$ where $E(A_1' S_p) = (3\sigma)/\sqrt{n}$, and $A_1' = 3/(c_2''\sqrt{n})$. For large k, it is expected that $s_p \approx s_p^*$ and both sets of control limits should be approximately equal. Let t_ν and $t_{\alpha,\nu}$ denote Student's t with ν degrees of freedom and the upper α percentile point, respectively. For $h = 1,2,\ldots,k, (\bar{X}^{(h)}-\bar{\bar{X}}) \sim N[0,(k-1)\sigma^2/kn]$, and $k(n-1)S_p^2/\sigma^2 \sim \chi^2[k(n-1)]$. Thus, $\sqrt{(kn)/(k-1)}\ (\bar{X}^{(h)}-\bar{\bar{X}})/S_p \sim t_{k(n-1)}$, and $P[\bar{\bar{X}}-t_{\alpha/2,k(n-1)}S_p\sqrt{(k-1)/kn} \le \bar{X}^{(h)} \le \bar{\bar{X}}+t_{\alpha/2,k(n-1)}S_p\sqrt{(k-1)/kn}] = 1 - \alpha$. Thus, the control chart constant A_1' may be thought of as corresponding to $t_{\alpha/2,k(n-1)}\sqrt{(k-1)/kn}$. Furthermore, since $A_1' = 3/(c_2''\sqrt{n})$, then $3/c_2''$ may be thought of as corresponding to $t_{\alpha/2,k(n-1)}\sqrt{1-k^{-1}}$. For p = 1, it is customary to choose $\alpha = .0027$ for theoretical control charts for the mean. Let the same α-level be chosen for these empirical charts. Thus, $t_{\alpha/2,k(n-1)} = t_{.00135,k(n-1)}$. When the control limits are written using percentage points of the t, the dependency of the limits upon the number of subgroups becomes very evident, a point which is frequently overlooked, since the control chart constants c_2 and c_2' are found using only n and not k. However, even in these cases, the control limits are indirectly dependent on k through $\bar{\bar{x}}$ and s_p^*. One final observation is that, regardless of p, it is meaningless to consider only one subgroup since $\bar{\underset{\sim}{x}}^{(1)} = \bar{\bar{\underset{\sim}{x}}}$.

In comparing $((kn)/(k-1))(\bar{\underset{\sim}{x}}^{(h)}-\bar{\bar{\underset{\sim}{x}}})'S_p^{-1}(\bar{\underset{\sim}{x}}^{(h)}-\bar{\bar{\underset{\sim}{x}}})$ with the percentage point $(k(n-1)p/(kn-k-p+1))F(p,kn-k-p+1,\alpha)$, the conclusion may have been that the process was out of control for the h*th* sample. The use of simultaneous techniques again permits the decision maker to determine those components responsible for the rejection.

Theorem 5.2. Let $X^{(1)},X^{(2)},\ldots,X^{(k)}$ be k $(n\times p)$ data matrices from $\underset{\sim}{X}$, where $\underset{\sim}{X}\sim N(\underset{\sim}{\mu},\Sigma)$ with both $\underset{\sim}{\mu}$ and Σ unknown. Then, for h = 1,2,...,k,

$$P[(\underset{\sim}{a}'\bar{\underset{\sim}{X}}^{(h)}-\underset{\sim}{a}'\bar{\bar{\underset{\sim}{X}}}) \le c(\underset{\sim}{a}'S_p\underset{\sim}{a})^{1/2} \text{ for all nonnull } \underset{\sim}{a}] = 1-\alpha, \qquad (19)$$

where $c = \sqrt{F(p,kn-k-p+1,\alpha)(k-1)k(n-1)p/((kn)(kn-k-p+1))}$.

Proof. $P[(kn/(k-1))(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})\le(k(n-1)p/(kn-k-p+1))F_\alpha]$ $= P[(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})\le c^2] = 1-\alpha$, where F_α denotes $F(p,kn-k-p+1,\alpha)$. Anderson (3) shows that S_p is positive definite with probability one. Thus, there exists a nonsingular matrix C such that $S_p = CC'$. Let $\underset{\sim}{y} = C^{-1}\bar{\underset{\sim}{x}}^{(h)}$. Then the inequality $(\bar{\underset{\sim}{x}}^{(h)}-\bar{\bar{\underset{\sim}{x}}})'S_p^{-1}(\bar{\underset{\sim}{x}}^{(h)}-\bar{\bar{\underset{\sim}{x}}}) \le c^2$ becomes $(\underset{\sim}{y}-C^{-1}\bar{\bar{\underset{\sim}{x}}})'(\underset{\sim}{y}-C^{-1}\bar{\bar{\underset{\sim}{x}}}) \le c^2$, or $\|\underset{\sim}{y}-C^{-1}\bar{\bar{\underset{\sim}{x}}}\|^2 \le c^2$. By Lemma 4.3, $\operatorname{mod}(\underset{\sim}{\ell}'\underset{\sim}{y}-\underset{\sim}{\ell}'C^{-1}\bar{\bar{\underset{\sim}{x}}}) \le c\|\underset{\sim}{\ell}\|$, for all $\underset{\sim}{\ell}$. Let $\underset{\sim}{a} = (C^{-1})'\underset{\sim}{\ell}$. Thus, the inequality becomes $\operatorname{mod}(\underset{\sim}{a}'\bar{\underset{\sim}{x}}^{(h)}-\underset{\sim}{a}'\bar{\bar{\underset{\sim}{x}}}) \le c(\underset{\sim}{a}'S_p\underset{\sim}{a})^{1/2}$ for all $\underset{\sim}{a}$, and Equation (19) follows. ||

Since the decision maker is interested in only a finite number of $\underset{\sim}{a}$, the confidence coefficient is at least $1-\alpha$. Specifically, if the h*th* sample plots out of control, the decision maker successively lets $\underset{\sim}{a}$ become $\underset{\sim}{e}_j$, $j = 1,2,\ldots,p$. Then, $\underset{\sim}{a}'S_p\underset{\sim}{a} = s_{pj}^2$, $j = 1,2,\ldots,p$, where s_{pj}^2 denotes the pooled sample variance for the j*th* variate. If the interval

$[-cs_{pj}, cs_{pj}]$ does not contain $\bar{x}_j^{(h)} - \bar{\bar{x}}_j$, then the decision maker would seek assignable causes of variation for the j*th* variate.

The Bonferroni technique, which follows, can also be used.

Theorem 5.3. Let $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$ be k $(n \times p)$ data matrices from $\underset{\sim}{X}$, where $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma)$ with both $\underset{\sim}{\mu}$ and Σ unknown. Let $\bar{\underset{\sim}{X}}^{(h)}$, $\bar{\bar{\underset{\sim}{X}}}$, and S_p be defined as previously. For $j = 1,2,\ldots,p$, let A_j denote the event that

$$\left(\bar{\bar{X}}_j - t_{\alpha/2p,k(n-1)} S_{pj} \sqrt{(k-1)/(kn)},\ \bar{\bar{X}}_j + t_{\alpha/2p,k(n-1)} S_{pj} \sqrt{(k-1)/(kn)}\right)$$

covers $\bar{X}_j^{(h)}$, where h is fixed and $\bar{\bar{X}}_j$ and S_{pj} denote the grand mean and pooled standard deviation, respectively, for the j*th* variate. Then $P\left(\bigcap_{j=1}^{p} A_j\right) \geq 1 - \alpha$, and the Bonferroni interval estimates are given by

$$\bar{\bar{x}}_j \pm t_{\alpha/2p,k(n-1)} s_{pj} \sqrt{(k-1)/(kn)}$$

Proof. The proof follows directly from Bonferroni's inequality which asserts that $P\left(\bigcap_{j=1}^{p} A_j\right) \geq 1 - \sum_{j=1}^{p} P(A_j^c) = 1 - \alpha$. ‖

For a coordinate by coordinate comparison on $\bar{\underset{\sim}{X}}^{(h)}$, the decision maker can easily determine which intervals are shorter by comparing $\sqrt{F(p,kn-k-p+1,\alpha)k(n-1)p/(kn-k-p+1)}$ with $t_{\alpha/2p,k(n-1)}$.

In checking the control of each subgroup by using the test statistic $n(\bar{\underset{\sim}{X}}^{(h)} - \bar{\bar{\underset{\sim}{X}}})' S_p^{-1} (\bar{\underset{\sim}{X}}^{(h)} - \bar{\bar{\underset{\sim}{X}}})$, the power is easily determined for each $h = 1,2,\ldots,k$, assuming that $\underset{\sim}{\mu}^{(1)} = \ldots = \underset{\sim}{\mu}^{(h-1)} = \underset{\sim}{\mu}^{(h+1)} = \ldots = \underset{\sim}{\mu}^{(k)} = \underset{\sim}{\mu}_0$. For fixed h, let $\underset{\sim}{X}^{(h)} \sim N(\underset{\sim}{\mu}^{(h)}, \Sigma)$ and let $\underset{\sim}{X}^{(j)} \sim N(\underset{\sim}{\mu}_0, \Sigma)$ for $j \neq h$.

Let $\underset{\sim}{Y} = \bar{\underset{\sim}{X}}^{(h)} - \bar{\bar{\underset{\sim}{X}}}$. Then $\underset{\sim}{Y} \sim N\{(1-k^{-1})(\underset{\sim}{\mu}^{(h)}-\underset{\sim}{\mu}_0),(k-1)\Sigma/kn\}$. If $\underset{\sim}{U} = \sqrt{(kn)/(k-1)}\underset{\sim}{Y}$, then $\underset{\sim}{U} \sim N\{\sqrt{n(1-k^{-1})}(\underset{\sim}{\mu}^{(h)}-\underset{\sim}{\mu}_0),\Sigma\}$. Since $\underset{\sim}{U}$ and S_p are independent, it follows from Theorem 5.1 that

$$n(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}}) \sim c(k,n,p)F'(p,kn-k-p+1,\lambda),$$

where $\lambda = n(1-k^{-1})(\underset{\sim}{\mu}^{(h)}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}^{(h)}-\underset{\sim}{\mu}_0)$. Thus

$$\pi(\lambda) = P\{F'(p,kn-k-p+1,\lambda)>F(p,kn-k-p+1)|\underset{\sim}{\mu}^{(h)}\}.$$

It is reasonable to replace Σ^{-1} by S_p^{-1} in determining λ. Also note that this noncentrality parameter equals the noncentrality parameter of a $\chi^2(p)$ as $k \to \infty$. Since the joint distribution of the k test statistics $n(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})$ is not presently known, the power of testing the overall equality of the k means cannot be determined, and the power discussed above might be called the marginal power.

One final property should be noted regarding the test statistic of interest. Since the test statistic for the theoretical chart was $n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)$ and $E(\bar{\bar{\underset{\sim}{X}}}) = \underset{\sim}{\mu}_0$ and $E(S_p) = \Sigma$, the test statistic $n(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})$ was used for the empirical chart. Kshirsagar (49) states that if A is $w(A|\Sigma,p,\nu)$ then $E(A^{-1}) = (\nu-p-1)^{-1}\Sigma^{-1}$ for $(\nu-p-1) > 0$. Thus, although S_p is an unbiased estimator of Σ, S_p^{-1} is not an unbiased estimator of Σ^{-1}. In fact, $(kn-k-p-1)S_p^{-1}/(kn-k)$ is an unbiased estimator of Σ^{-1}, and one should find the distribution of $\{n(kn-k-p-1)/(kn-k)\}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})$. However, this statistic has the previously derived distribution. That is,

$$\left((kn-k-p-1)/(kn-k)\right)(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})\sim\left((kn-k-p-1)(k-1)/(kn)(kn-k-p+1)F\right.$$

where F denotes F(p,kn-k-p+1).

Viewing the control of the process as a hypothesis testing problem, one would set up the null hypothesis H_0: $\underset{\sim}{\mu}^{(1)} = \underset{\sim}{\mu}^{(2)} = \ldots = \underset{\sim}{\mu}^{(k)}$ against all possible alternatives. Specifically, Ω and ω are those subsets of Euclidean $(p^2+p+2kp)/2$ space such that

$$\Omega = \{(\underset{\sim}{\mu}^{(1)},\underset{\sim}{\mu}^{(2)},\ldots,\underset{\sim}{\mu}^{(k)},\Sigma)\text{: } -\infty<\underset{\sim}{\mu}^{(h)}<\infty,\ \Sigma \text{ is positive definite, } h=1,2,\ldots,k\}$$

and

$$\omega = \{(\underset{\sim}{\mu}^{(1)},\underset{\sim}{\mu}^{(2)},\ldots,\underset{\sim}{\mu}^{(k)},\Sigma)\text{: } -\infty<\underset{\sim}{\mu}^{(1)}=\underset{\sim}{\mu}^{(2)}=\ldots=\underset{\sim}{\mu}^{(k)}<\infty,\ \Sigma \text{ is positive definite}\},$$

where $-\infty < \underset{\sim}{\mu}^{(h)} < \infty$ implies that each component of the vector $\underset{\sim}{\mu}^{(h)}$ lies between $-\infty$ and $+\infty$. Let $L(\omega)$, $L(\hat{\omega})$ and $L(\Omega)$, $L(\hat{\Omega})$ denote the likelihood functions and their supremum for the parameter spaces ω and Ω, respectively. If $\underset{\sim}{\mu}$ denotes the common but unspecified value of $\underset{\sim}{\mu}^{(1)},\underset{\sim}{\mu}^{(2)},\ldots,\underset{\sim}{\mu}^{(k)}$, then

$$L(\omega) = \prod_{i=1}^{k}\left[(2\pi)^{-pn/2}|\Sigma^{-1}|^{n/2}\exp\left(-(1/2)\sum_{j=1}^{n}(\underset{\sim}{x}_j^{(i)}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_j^{(i)}-\underset{\sim}{\mu})\right)\right] \qquad (20)$$

and

$$L(\Omega) = \prod_{i=1}^{k}\left[(2\pi)^{-pn/2}|\Sigma^{-1}|^{n/2}\exp\left(-(1/2)\sum_{j=1}^{n}(\underset{\sim}{x}_j^{(i)}-\underset{\sim}{\mu}^{(i)})'\Sigma^{-1}(\underset{\sim}{x}_j^{(i)}-\underset{\sim}{\mu}^{(i)})\right)\right] \qquad (21)$$

The logarithms of Equations (20) and (21) are the sums of k terms which can be maximized separately because of the independence of the k subgroups. By using a procedure similar to that used in the proof of Theorem 3.1, one obtains

$$\hat{\underset{\sim}{\mu}}_{\omega} = \bar{\bar{\underset{\sim}{x}}} = (kn)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n} \underset{\sim}{x}_j^{(i)}, \tag{22}$$

$$\hat{\Sigma}_{\omega} = (kn)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n} (\underset{\sim}{x}_j^{(i)} - \bar{\bar{\underset{\sim}{x}}})(\underset{\sim}{x}_j^{(i)} - \bar{\bar{\underset{\sim}{x}}})' \tag{23}$$

$$\hat{\underset{\sim}{\mu}}_{\Omega}^{(i)} = \bar{\underset{\sim}{x}}^{(i)} = n^{-1} \sum_{j=1}^{n} \underset{\sim}{x}_j^{(i)}, \quad i = 1,2,\ldots,k, \tag{24}$$

and

$$\hat{\Sigma}_{\Omega} = (kn)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n} (\underset{\sim}{x}_j^{(i)} - \bar{\underset{\sim}{x}}^{(i)})(\underset{\sim}{x}_j^{(i)} - \bar{\underset{\sim}{x}}^{(i)})' \tag{25}$$

$$= (kn)^{-1}C. \tag{26}$$

Since $\sum_{j=1}^{n} (\underset{\sim}{x}_j^{(i)} - \bar{\bar{\underset{\sim}{x}}})(\underset{\sim}{x}_j^{(i)} - \bar{\bar{\underset{\sim}{x}}})' = \sum_{j=1}^{n} (\underset{\sim}{x}_j^{(i)} - \bar{\underset{\sim}{x}}^{(i)})(\underset{\sim}{x}_j^{(i)} - \bar{\underset{\sim}{x}}^{(i)})' + n(\bar{\underset{\sim}{x}}^{(i)} - \bar{\bar{\underset{\sim}{x}}})(\bar{\underset{\sim}{x}}^{(i)} - \bar{\bar{\underset{\sim}{x}}})'$, Equation (23) can be written as $\hat{\Sigma}_{\omega} = (kn)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n} (\underset{\sim}{x}_j^{(i)} - \bar{\underset{\sim}{x}}^{(i)})(\underset{\sim}{x}_j^{(i)} - \bar{\underset{\sim}{x}}^{(i)})' + (kn)^{-1} \sum_{i=1}^{k} n(\bar{\underset{\sim}{x}}^{(i)} - \bar{\bar{\underset{\sim}{x}}})(\bar{\underset{\sim}{x}}^{(i)} - \bar{\bar{\underset{\sim}{x}}})' = (kn)^{-1}(C+D)$, where $D = n \sum_{i=1}^{k} (\bar{\underset{\sim}{x}}^{(i)} - \bar{\bar{\underset{\sim}{x}}})(\bar{\underset{\sim}{x}}^{(i)} - \bar{\bar{\underset{\sim}{x}}})'$. Thus,

$$L(\hat{\omega}) = \left[\frac{kn}{(2\pi)^p |C+D|}\right]^{kn/2} e^{-kn/2},$$

and

$$L(\hat{\Omega}) = \left[\frac{kn}{(2\pi)^p|C|}\right]^{kn/2} e^{-kn/2},$$

and

$$\lambda(X^{(1)},\ldots,X^{(k)}) = \left[\frac{|C|}{|C+D|}\right]^{kn/2}. \tag{27}$$

This result was first derived by Wilks (70). It is customary to denote the 2/kn power of Equation (27) by Λ, and this is called Wilk's Λ criterion. Thus,

$$\Lambda = |C||C+D|^{-1}, \tag{28}$$

and the form of the critical region is given by

$$w = \{(X^{(1)},\ldots,X^{(k)}):\ |C||C+D|^{-1}<k\}, \tag{29}$$

where k needs to be determined such that the probability of type I error is α. Wilk's Λ criterion arises in many areas of multivariate statistics and has received considerable attention in the past ten years.

When p = 1 and the null hypothesis is true, $\Lambda = (1+q_1/q_2)^{-1}$ where $q_1 = n\sum_{i=1}^{k}(\bar{x}_i-\bar{\bar{x}})^2$ and $q_2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2$. However, $(kn-k)Q_1/(k-1)Q_2$ is distributed as an $F(k-1,k(n-1))$. Consequently,

$\Lambda = \left(1+\left((k-1)/(kn-k)\right)F\right)^{-1}$, where F denotes F(k-1,kn-k). Equivalently, $F = (kn-k)(1-\Lambda)\left((k-1)\Lambda\right)^{-1}$, and one would reject H_0 whenever F exceeded $F(k-1,kn-k,\alpha)$, a result stated earlier.

When $p \geq 2$ and the null hypothesis is true, Wilks and others have solved the distributional problem. Let ν_1 denote the d.f. of the total sums of squares and products matrix C + D; let ν_2 denote the d.f. of the sums of squares and products matrix D due to hypothesis; let p denote the number of variables. To denote this, the test statistic Λ is sometimes written $\Lambda(p,\nu_1,\nu_2)$. For H_0: $\underset{\sim}{\mu}^{(1)} = \underset{\sim}{\mu}^{(2)} = \ldots = \underset{\sim}{\mu}^{(k)}$, $\nu_1 = kn - 1$ and $\nu_2 = k - 1$. If X denotes the standard beta random variable with parameters α and β, then $f_X(x) = \left(B(\alpha,\beta)\right)^{-1}x^{\alpha-1}(1-x)^{\beta-1}$, $0 \leq x \leq 1$, and $E(X^r) = [\Gamma(\alpha+\beta)\Gamma(\alpha+r)]/[\Gamma(\alpha)\Gamma(\alpha+\beta+r)]$. Wilks (70) showed that $E(\Lambda^r) = \prod_{h=1}^{p} E(X_h^r)$, where the X_h are independent standard beta random variables with $\alpha_h = (1/2)(\nu_1-\nu_2+1-h)$ and $\beta_h = \nu_2/2$. Since the range of each X_h is finite, the moments determine the distribution uniquely. By a comparison of moments, some particular cases are presented below in Table 7. Bartlett (5) obtained the characteristic function of $W = -m \ln \Lambda$ from $E(\Lambda^r)$, and derived a first approximation by stating that W is approximately distributed as $\chi^2(p\nu_2)$ where $m = \nu_1 - (1/2)(p+\nu_2+1)$. Rao (57) obtained a better first approximation by using

$$R = \frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}} \frac{ms + 2\lambda}{p\nu_2} ,$$

where R is approximately distributed as an $F(p\nu_2, ms-2\lambda)$ with

$s = \sqrt{(p^2\nu_2^2-4)/(p^2+\nu_2^2-5)}$ and $\lambda = -(p\nu_2-2)/4$. Box (9) extended Bartlett's result and gave a higher order approximation.

Table 7. The Equivalence of Functions of Λ and F for Certain Values of p and ν_2

$\nu_2 = 1$, any p	$\frac{1-\Lambda}{\Lambda}\,\frac{\nu_1-p}{p} = F(p,\nu_1-p)$
$\nu_2 = 2$, any p	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\,\frac{\nu_1-p-1}{p} = F\big(2p,2(\nu_1-p-1)\big)$
$p = 1$, any ν_2	$\frac{1-\Lambda}{\Lambda}\,\frac{\nu_1-\nu_2}{\nu_2} = F(\nu_2,\nu_1-\nu_2)$
$p = 2$, any ν_2	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\,\frac{\nu_1-\nu_2-1}{\nu_2} = F\big(2\nu_2,2(\nu_1-\nu_2-1)\big)$

Schatzoff (61) has prepared tables from which the exact percentage points of W can be obtained, which are denoted by W^*. The tables have values of multiplying factors c as their entries such that $W^* = c\chi^2(p\nu_2,\alpha)$, and the null hypothesis is rejected at the α level if $W > W^*$. For (p,ν_2) = (3,4), (4,4), (5,4), (6,4), (7,4), (8,4), (9,4), (10,4), (3,6), (4,6), (5,6), (6,6), (7,6), (8,6), (9,6), (10,6), (3,8), (5,8), (7,8), (8,8), (3,10), (5,10), (7,10), and α = .005, .010, .025, .050, .100, one can find values of c for M = 1(1)10, 12(2),20,24, 30,40,60,120,∞, where $M = \nu_1 - \nu_2 - p + 1$. Anderson (3) has shown that $\Lambda(p,\nu_2,\nu_1-\nu_2) = \Lambda(\nu_2,p,\nu_1-p)$. This enables one to use Schatzoff's tables when ν_2 is odd and p is even. M is invariant under this change.

Note that the tables only contain values of multiplying factors for ν_2 even or p even. When p and ν_2 are both odd, approximations to c can be obtained by linear interpolation. The power of the test is not discussed here since it involves characteristic roots. However, Kshirsagar (49) does present a detailed development for the null hypothesis under consideration. Miller (53) presents a simultaneous technique.

2. Tests of Significance and Hotelling's T^2

In this chapter, the population parameters governing the repetitive process were unknown, and unbiased estimates were obtained for them. The process was judged to be in control relative to $\bar{\bar{\underset{\sim}{x}}}$. However, the process may not be in control at some target value of the population mean, denoted by $\underset{\sim}{\mu}_0$. Bowker and Lieberman (8) describe this for the case of one quality characteristic as follows:

> For example, suppose that an item is being produced according to a probability distribution which has a mean, $\bar{x}'$, equal to 20. If the aimed-at-value is $\bar{x}' = 25$, the control chart based on $\bar{x}' = 25$ will exhibit a lack of control. . . . with the present machine setting, the process is in control at $\bar{x}' = 20$. The term "state of control" should be interpreted in this light.

Since $\bar{\bar{\underset{\sim}{x}}}$ should be "close" to the actual population mean, the control of the process was determined relative to $\bar{\bar{\underset{\sim}{x}}}$ and not some aimed-at-value $\underset{\sim}{\mu}_0$. When a state of statistical control has been obtained at $\bar{\bar{\underset{\sim}{x}}}$, an entirely new problem arises in changing the process to a new state of control using some aimed-at-value or determining the control of the process at this aimed-at-value. In these situations, the aimed-at-value is denoted by $\underset{\sim}{\mu}_0$. Again, Σ is assumed to be unknown.

The first situation that may arise is when a random sample of size m has been obtained in some period, and it is necessary to determine whether the process is in control relative to $\underset{\sim}{\mu}_0$ in this period. This can be viewed as a hypothesis testing problem, viz.,

$$H_0: \underset{\sim}{\mu} = \underset{\sim}{\mu}_0 \quad \text{versus} \quad H_1: \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$$

with Σ unknown. To test $H_0: \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ versus $H_1: \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ by the likelihood ratio principle, based on a random sample of size m,

$$\lambda(X) = \frac{\sup\{L(X|\underset{\sim}{\mu},\Sigma): (\underset{\sim}{\mu},\Sigma)\epsilon\omega\}}{\sup\{L(X|\underset{\sim}{\mu},\Sigma): (\underset{\sim}{\mu},\Sigma)\epsilon\Omega\}}, \tag{30}$$

where Ω and ω are those subsets of Euclidean $(p^2+3p)/2$ space such that

$$\omega = \{(\underset{\sim}{\mu},\Sigma): \underset{\sim}{\mu} = \underset{\sim}{\mu}_0,\ \Sigma \text{ is positive definite}\},$$

$$\Omega = \{(\underset{\sim}{\mu},\Sigma): -\infty < \underset{\sim}{\mu} < \infty,\ \Sigma \text{ is positive definite}\}.$$

Let $L(\hat{\omega})$ and $L(\hat{\Omega})$ denote the numerator and denominator, respectively, of Equation (30). Using a procedure similar to that used in the proof of Theorem 3.1, one obtains

$$\hat{\underset{\sim}{\mu}}_\omega = \underset{\sim}{\mu}_0,$$

$$\hat{\Sigma}_\omega = (1/m) \sum_{h=1}^{m} (\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)(\underset{\sim}{x}_h-\underset{\sim}{\mu}_0)'.$$

From Theorem 3.1, it immediately follows that

$$\hat{\underset{\sim}{\mu}}_\Omega = \bar{\underset{\sim}{x}},$$

$$\hat{\Sigma}_\Omega = (1/m) \sum_{h=1}^{m} (\underset{\sim}{x}_h-\bar{\underset{\sim}{x}})(\underset{\sim}{x}_h-\bar{\underset{\sim}{x}})' = (1/m)A.$$

Thus,

$$\lambda(X) = \frac{|\hat{\Sigma}_\omega|^{-m/2}}{|\hat{\Sigma}_\Omega|^{-m/2}} = \frac{|\hat{\Sigma}_\Omega|^{m/2}}{|\hat{\Sigma}_\omega|^{m/2}} \tag{31}$$

Now $\hat{\Sigma}_\omega = (1/m)A + (\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'$. Thus

$$\lambda^{2/m} = \frac{|A|}{|A+m(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'|}, \tag{32}$$

where $\lambda^{2/m}$ denotes $\{\lambda(X)\}^{2/m}$. Note $\lambda^{2/m}$ is a special case of Wilk's Λ criterion. Since $A + m(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)' = A\{I+mA^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'\}$, Equation (32) becomes

$$\lambda^{2/m} = |I+mA^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'|^{-1}. \tag{33}$$

By Lemma 2.5, the determinant of the matrix

$$\left[\begin{array}{c|c} 1 & -\sqrt{m}\,(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)' \\ \hline \sqrt{m}\,A^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0) & I_p \end{array}\right]$$

equals the reciprocal of Equation (33), which also equals $|1+m(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'A^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)|$. But, the determinant of a scalar equals the scalar. Thus, $1 + m(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'A^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0) = |I+mA^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'|$, and

$$\lambda^{2/m} = \left(1+T^2(m-1)^{-1}\right)^{-1}, \tag{34}$$

where $T^2 = m(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'S^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)$. The form of the critical region is given by

$$w = \{X\colon\ m(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'S^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)>c\}. \tag{35}$$

This result is presented in Anderson (3). A solution to the general distributional problem can be formulated using Theorem 5.1.

Theorem 5.4. Let $\underset{\sim}{X}_1,\underset{\sim}{X}_2,\ldots,\underset{\sim}{X}_m$ be a random sample from $\underset{\sim}{X}$, where $\underset{\sim}{X}\sim N(\underset{\sim}{\mu},\Sigma)$. Then $m(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'S^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)$ is distributed as $\left(p(m-1)/(m-p)\right)F'(p,m-p,\lambda)$, where $\lambda = m(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$.

Proof. From Theorem 3.2, $\bar{\underset{\sim}{X}}\sim N(\underset{\sim}{\mu},\Sigma/m)$. Thus, $\sqrt{m}\,(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)$ is normally distributed with mean vector $\sqrt{m}\,(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$ and covariance matrix Σ. From Theorem 3.2, $(m-1)S$ is independently distributed as $\sum_{h=1}^{m-1} \underset{\sim}{Y}_h\underset{\sim}{Y}_h'$ with the

$\underset{\sim}{Y}_h$ independent and $\underset{\sim}{Y}_h \sim N(\underset{\sim}{0},\Sigma)$, $h = 1,2,\ldots,m-1$. Then, by Theorem 5.1, $\{T^2/(m-1)\}\{(m-p)/p\}$ is distributed as $F'(p,m-p,\lambda)$ where $T^2 = m(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'S^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)$. ‖

When $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}_0,\Sigma)$, then $\lambda = 0$, and $\{T^2/(m-1)\}\{(m-p)/p\}$ is distributed as a central F random variable. The distribution of T^2 under the null hypothesis was first derived by Hotelling (35) and is called Hotelling's T^2 distribution. Hence, the critical constant c in Equation (35) is the modified percentage point $\{p(m-1)/(m-p)\}F(p,m-p,\alpha)$ such that

$$P\{m(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'S^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)>\{p(m-1)/(m-p)\}F(p,m-p,\alpha)\,|\,\underset{\sim}{\mu}=\underset{\sim}{\mu}_0\} = \alpha.$$

In testing $H_0\colon \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ versus $H_1\colon \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$, the decision maker would take a random sample of size m, compute $m(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)'S^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}_0)$, and determine whether this quantity exceeds $\{p(m-1)/(m-p)\}F(p,m-p,\alpha)$, where $F(p,m-p,\alpha)$ is obtained from tables of the F distribution.

The power of the test $H_0\colon \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs. $H_1\colon \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ is denoted $\pi(\lambda)$, where

$$\pi(\lambda) = P\{m(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'S^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)>\{p(m-1)/(m-p)\}F(p,m-p,\alpha)\,|\,\underset{\sim}{\mu}\}.$$

Equivalently, $\pi(\lambda) = P\{F'(p,m-p,\lambda)>F(p,m-p,\alpha)\,|\,\underset{\sim}{\mu}\}$. Lachenbruch (50) gives values of these probabilities for $\alpha = .01, .05$, $\nu_1 = 1(1)10, 12(2)16, 20, 24, 30, 40, 50, 75$, $\nu_2 = 2(2)20, 24(4)36, 40(10)80$, and $\sqrt{\lambda/(\nu_1+1)} = 1(.5)3, 4(1)8$. Tiku (67) has published tables for $\alpha = .005, .01, .025, .05$, $\nu_1 = 1(1)10, 12$, $\nu_2 = 2(2)30, 40, 60, 120, \infty$, and

$\sqrt{\lambda/(\nu_1+1)} = .5(.5)3.0$. However, the easiest computation of the power is by the use of the Pearson and Hartley (56) charts. The charts are expressed in terms of a quantity ϕ, where $\phi = \sqrt{\lambda/(p+1)}$. An abridged version of these tables is contained in Morrison (55). Unfortunately, λ depends on Σ which is unknown. One reasonable technique is to replace Σ by its unbiased estimate. Since $F'(\nu_1,\infty,\lambda) = \chi'^2(\nu_1,\lambda)/\nu_1$, these charts can also be used to obtain the power of the test on the means when Σ is known (Chapter IV) by letting $\nu_1 = p$, $\nu_2 = \infty$, and $\lambda = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$. For Σ known or unknown, a paradox arises involving the power of the test. This paradox and its resolutions are treated in Chapter VI.

In testing H_0: $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs. H_1: $\underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ with Σ unknown, confidence intervals for the p coordinate means or other parametric functions of $\underset{\sim}{\mu}$ can be obtained based on the presentation in Chapter IV. Miller (53) gives a detailed account of their development.

The first method is based on Bonferroni's inequality: $P\left(\bigcap_{h=1}^{p} A_h\right) \geq 1 - \sum_{h=1}^{p} P(A_h^c)$. If $X\sim N(\mu,\sigma^2)$ with both μ and σ^2 unknown, then the two statistics $\underline{\mu} = \bar{X} - t_{\gamma/2,m-1}(S/\sqrt{m})$ and $\bar{\mu} = \bar{X} + t_{\gamma/2,m-1}(S/\sqrt{m})$ constitute a $100(1-\gamma)\%$ confidence interval for μ based on a random sample of size m. Since $\underset{\sim}{X}\sim N(\underset{\sim}{\mu},\Sigma)$ and $X_h\sim N(\mu_h,\sigma_h^2)$, $h = 1,\ldots,p$, then $\bar{X}_h\sim N(\mu_h,\sigma_h^2/m)$ and $(m-1)S_h^2/\sigma_h^2\sim\chi^2(m-1)$, where $\bar{X}_h = (1/m)\sum_{j=1}^{m} X_{hj}$ and $S_h^2 = (m-1)^{-1}\sum_{j=1}^{m}(X_{hj}-\bar{X}_h)^2$. Let $\underline{\mu}_h = \bar{X}_h - t_{\alpha/2p,m-1}(S_h/\sqrt{m})$ and $\bar{\mu}_h = \bar{X}_h + t_{\alpha/2p,m-1}(S_h/\sqrt{m})$, and let A_h denote the event that $(\underline{\mu}_h,\bar{\mu}_h)$ covers μ_h. Then $P\left(\bigcap_{h=1}^{p} A_h\right) \geq 1 - \alpha$, and the Bonferroni confidence interval estimates are given by

$$\bar{x}_h \pm t_{\alpha/2p,m-1}(s_h/\sqrt{m}) \tag{36}$$

The percentage points $t_{\alpha/2p,m-1}$ have been tabulated by Dunn (20) for α = .05, .01, p = 2(1)9, 10(5)50,100,250, ν = 5, 7, 10, 12, 15, 20, 24, 30, 40, 60, 120, ∞, where ν = m-1. A nomogram has been prepared by James-Levy (43) which permits a rapid determination of these points. For any k linear combinations, the Bonferroni intervals can be found by introducing the appropriate $\underset{\sim}{a}$ in

$$\underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\bar{\underset{\sim}{x}} \pm (\sqrt{m}\ t_{\alpha/2k,m-1})(\underset{\sim}{a}'S\underset{\sim}{a})^{1/2}$$

The second method is based on Scheffe's F technique and is presented in the following theorem. A proof is not presented since it closely parallels the proof of Theorem 4.3. However, it must be assumed that S is positive definite so that S = CC'.

<u>Theorem 5.5</u>. Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_m$ be a random sample from $\underset{\sim}{X}$, where $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma)$ with both $\underset{\sim}{\mu}$ and Σ unknown. Then

$$P[\{\underset{\sim}{a}'(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})\}^2 \le c^2(\underset{\sim}{a}'S\underset{\sim}{a}) \text{ for all nonnull } \underset{\sim}{a}] = 1 - \alpha, \tag{37}$$

where $c^2 = \{p(m-1)/(m-p)m\}F(p,m-p,\alpha)$.

Since the decision maker is interested in only a finite number of $\underset{\sim}{a}$, the confidence coefficient is at least $1 - \alpha$. An immediate consequence of Equation (37) and Lemma 4.5 is that the confidence interval estimates for μ_h, $h = 1,\ldots,p$ are given by

$$\bar{x}_h \pm \left(p(m-1)/(m-p)m\right)^{1/2}\left(F(p,m-p,\alpha)\right)^{1/2} s_h \tag{38}$$

Other parametric functions of $\underset{\sim}{\mu}$ can be obtained by varying the elements of $\underset{\sim}{a}$.

The only difference between the Bonferroni intervals and the Scheffe intervals are the constants $t_{\alpha/2p,m-1}$ and $\left[\left(p(m-1)/(m-p)F(p,m-p,\alpha)\right)\right]^{1/2}$, respectively. The decision maker can easily determine which constant is smaller.

As mentioned at the outset of this section, the Hotelling T^2 presented herein is applicable to situations where a random sample of size m has been obtained in some period and it was desired to determine process control relative to $\underset{\sim}{\mu}_0$ in this period. There are several cases where this occurs. After having obtained k rational subgroups of size n, the decision maker might be interested in determining whether the process is in control at some specified value $\underset{\sim}{\mu}_0$; in this case, $m = kn$. Another case occurs when the decision maker has obtained a random sample of m vectors and wishes to determine process control relative to $\underset{\sim}{\mu}_0$; for example, this situation could arise when production commences. Since neither of these cases are repetitive types of procedures, control charts would not be established. However, the control chart concept could arise in the case where a random sample of size m is obtained on successive occasions and the sample variance-covariance matrix and the sample mean vector are computed each time; in this case a control chart could be established with control limit given by $[p(m-1)/(m-p)]F(p,m-p,\alpha)$.

Another use of Hotelling's T^2 would be in the control of future observations where the sample variance-covariance matrix is determined from past data. Furthermore, the state of control is the aimed-at value denoted by $\underset{\sim}{\mu}_0$. The distributional problem is treated in the following theorem.

<u>Theorem 5.6.</u> Let $\underset{\sim}{X} \sim N(\underset{\sim}{\mu},\Sigma)$. Based on an initial random sample of size m, $S_m = (m-1)^{-1} \sum_{h=1}^{m} (\underset{\sim}{X}_{h1}-\bar{\underset{\sim}{X}}_m)(\underset{\sim}{X}_{h1}-\bar{\underset{\sim}{X}}_m)'$. Let $\underset{\sim}{X}_{12},\underset{\sim}{X}_{22},\ldots,\underset{\sim}{X}_{n2}$ denote the elements of a successive random sample, where $\bar{\underset{\sim}{X}}_n = (1/n) \sum_{h=1}^{n} \underset{\sim}{X}_{h2}$. Then $n(\bar{\underset{\sim}{X}}_n-\underset{\sim}{\mu}_0)'S_m^{-1}(\bar{\underset{\sim}{X}}_n-\underset{\sim}{\mu}_0)$ is distributed as $[p(m-1)/(m-p)]F'(p,m-p,\lambda)$, where $\lambda = m(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$.

Proof. From Theorem 3.2, $\bar{\underset{\sim}{X}}_n \sim N(\underset{\sim}{\mu},\Sigma/n)$. Thus, $\sqrt{n}\,(\bar{\underset{\sim}{X}}_n-\underset{\sim}{\mu}_0)$ is distributed as $N[\sqrt{n}\,(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0),\Sigma]$. In addition, $(m-1)S_m$ is independently distributed as $\sum_{h=1}^{m-1} \underset{\sim}{Y}_h\underset{\sim}{Y}_h'$ with the $\underset{\sim}{Y}_h$ independent and $\underset{\sim}{Y}_h \sim N(\underset{\sim}{0},\Sigma)$, $h = 1,2,\ldots,m-1$. Then, by Theorem 5.1, $[T^2/(m-1)][(m-p)/p]$ is distributed as $F'(p,m-p,\lambda)$ where $T^2 = n(\bar{\underset{\sim}{X}}_n-\underset{\sim}{\mu}_0)'S_m^{-1}(\bar{\underset{\sim}{X}}_n-\underset{\sim}{\mu}_0)$. ‖

Thus, the decision maker would compute $S_m,\bar{\underset{\sim}{x}}_n$, and determine whether $n(\bar{\underset{\sim}{x}}_n-\underset{\sim}{\mu}_0)'S_m^{-1}(\bar{\underset{\sim}{x}}_n-\underset{\sim}{\mu}_0)$ exceeds $[p(m-1)/(m-p)]F(p,m-p,\alpha)$.

This process could be repeated for additional random samples of size n, and a control chart could be established. Note that n could equal one. Furthermore,

$$\pi(\lambda) = P\left(n(\bar{\underset{\sim}{X}}_n-\underset{\sim}{\mu}_0)'S_m^{-1}(\bar{\underset{\sim}{X}}_n-\underset{\sim}{\mu}_0)\geq\left(p(m-1)/(m-p)\right)F(p,m-p,\alpha)\,\middle|\,\underset{\sim}{\mu}\right),$$

which can easily be determined for each new random sample of size n. Also, confidence intervals for any parametric function of $\underset{\sim}{\mu}$ are easily obtained using either the Bonferroni or Scheffe techniques. A modification of this control procedure is suggested. For $h = 1,2,\ldots,$ let n_h denote the sizes of the additional random samples. If the test statistic based on the n_1 data vectors does not indicate lack of control, then the statistic for the next n_2 data vectors becomes $n_2(\bar{\underset{\sim}{x}}_{n_2}-\underset{\sim}{\mu}_0)'S^{-1}_{m+n_1}(\bar{\underset{\sim}{x}}_{n_2}-\underset{\sim}{\mu}_0)$, where $S_{m+n_1} = \left((m-1)S_m+(n_1-1)S_n\right)/(m+n_1-2)$ and the critical constant becomes $\left(p(m+n_1-2)/(m+n_1-1-p)\right)F(p,m+n_1-1-p,\alpha)$. In view of the updating of the sample variance-covariance matrix, it would be of no benefit to establish a control chart limit since it would change for each successive sample. This updating takes on added importance if a separate control chart is being maintained for S_{n_h}. Furthermore, this updating could take place less frequently than after each sample of size n_h.

The previous uses of Hotelling's T^2 were to determine the control of a process relative to $\underset{\sim}{\mu}_0$. A modification of this is the determination of control relative to some sample mean where both this sample mean and the sample variance-covariance matrix are obtained from past data. The distributional problem is treated in the following theorem.

Theorem 5.7. Let $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma)$. Based on an initial random sample of size m, $\bar{\underset{\sim}{X}}_m = (1/m) \sum_{h=1}^{m} \underset{\sim}{X}_{h1}$, and $S_m = (m-1)^{-1} \sum_{h=1}^{m} (\underset{\sim}{X}_{h1} - \bar{\underset{\sim}{X}}_m)(\underset{\sim}{X}_{h1} - \bar{\underset{\sim}{X}}_m)'$. Let $\underset{\sim}{X}_{12}, \underset{\sim}{X}_{22}, \ldots, \underset{\sim}{X}_{n2}$ denote the elements of a successive random sample, where $\bar{\underset{\sim}{X}}_n = (1/n) \sum_{h=1}^{n} \underset{\sim}{X}_{h2}$. Then $n(\bar{\underset{\sim}{X}}_n - \bar{\underset{\sim}{X}}_m)' S_m^{-1} (\bar{\underset{\sim}{X}}_n - \bar{\underset{\sim}{X}}_m)$ is distributed as

$$\left[p(m+n)(m-1)/(m^2-mp)\right] F(p, m-p)$$

Proof. Since $\bar{\underset{\sim}{X}}_n \sim N(\underset{\sim}{\mu}, \Sigma/n)$ and $\bar{\underset{\sim}{X}}_m \sim N(\underset{\sim}{\mu}, \Sigma/m)$, then $\underset{\sim}{U} = \sqrt{(mn)/(m+n)}\, (\bar{\underset{\sim}{X}}_n - \bar{\underset{\sim}{X}}_m) \sim N(\underset{\sim}{0}, \Sigma)$. In addition, $(m-1)S_m$ is independently distributed as $\sum_{h=1}^{m-1} \underset{\sim}{Y}_h \underset{\sim}{Y}_h'$ with the $\underset{\sim}{Y}_h$ independent and $\underset{\sim}{Y}_h \sim N(\underset{\sim}{0}, \Sigma)$, $h = 1,2,\ldots,m-1$. Then, $\left[mn/(m+n)\right](\bar{\underset{\sim}{X}}_n - \bar{\underset{\sim}{X}}_m)' S_m^{-1} (\bar{\underset{\sim}{X}}_n - \bar{\underset{\sim}{X}}_m)$ is distributed as $\left[p(m-1)/(m-p)\right] F(p, m-p)$. Equivalently, $n(\bar{\underset{\sim}{X}}_n - \bar{\underset{\sim}{X}}_m)' S_m^{-1} (\bar{\underset{\sim}{X}}_n - \bar{\underset{\sim}{X}}_m)$ is distributed as $c(m,n,p)F(p,m-p)$ where $c(m,n,p) = \left[p(m+n)(m-1)/(m^2-mp)\right]$. ||

Thus, the decision maker would compute $\bar{\underset{\sim}{x}}_m$, S_m, $\bar{\underset{\sim}{x}}_n$, and determine whether $n(\bar{\underset{\sim}{x}}_n - \bar{\underset{\sim}{x}}_m)' S_m^{-1} (\bar{\underset{\sim}{x}}_n - \bar{\underset{\sim}{x}}_m)$ exceeds $c(m,n,p)F(p,m-p,\alpha)$. This process could be repeated for additional random samples of size n, and a control chart could be established. Note that n could equal one. Confidence intervals for any parametric function are easily obtained using either the Bonferroni or Scheffe techniques. Furthermore, either S_m or $\bar{x}_m$ could be updated prior to each additional random sample or less frequently.

The literature, including both the history and apparent errors, can now be reviewed. Jackson (40,41,42) has apparently incorrectly applied one form of Hotelling's T^2 distribution in two separate, but related papers. Using a "suitable base period of 75 successive analyses, during which no trouble had occurred . . .," he attempts to analyze these 75 past observation vectors for lack of control by using the test statistic T_J^2 for h = 1,2,...,75, where

$$T_J^2 = (\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})'S^{-1}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}}),$$

and $\bar{\underset{\sim}{X}}$ and S are based on the 75 observations. Jackson claims that $T_J^2 = \{p(n-1)/(n-2)\}F(p,n-2)$, where p = 2 and n = 75. Unfortunately, T_J^2 does not seem to follow Hotelling's T^2 distribution since $(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})$ and S^{-1} are not independent. The proof of this result is based on a necessary and sufficient condition presented by Rao (58), which is a multivariate analogue of Craig's (13) theorem for the independence of a linear and quadratic form. If $\underset{\sim}{X}_h \sim N(\underset{\sim}{\mu},\Sigma)$, h = 1,2,...,n, and the $\underset{\sim}{X}_h$ are independent and $X = [\underset{\sim}{X}_1,\underset{\sim}{X}_2,\ldots,\underset{\sim}{X}_n]$, then Rao proves that XBX' and $X\underset{\sim}{C}$ are independently distributed if and only if $\underset{\sim}{C}'B = \underset{\sim}{0}'$.

<u>Remark 5.2</u>. S and $(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})$ are not independently distributed, where $\underset{\sim}{X}_h \sim N(\underset{\sim}{\mu},\Sigma)$, for h = 1,2,...,n.

Proof. Since $S = (n-1)^{-1}\sum_{h=1}^{n}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})'$, S can be represented as $S = (n-1)^{-1}X(I-n^{-1}U)X'$, where U denotes the (n×n) unit matrix. Further, $\bar{\underset{\sim}{X}} = X(n^{-1}\underset{\sim}{j})$, where $\underset{\sim}{j}$ denotes the (n×1) unit vector. Thus, $(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}}) = X\underset{\sim}{e}_h - X(n^{-1}\underset{\sim}{j}) = X(\underset{\sim}{e}_h-n^{-1}\underset{\sim}{j})$, where $\underset{\sim}{e}_h = \delta_h\underset{\sim}{j}$. Since

$$(\underset{\sim}{e}_h-n^{-1}\underset{\sim}{j})'(I-n^{-1}U) = (-n^{-1},\ldots,-n^{-1},1-n^{-1},-n^{-1},\ldots,-n^{-1}) \neq \underset{\sim}{0}',$$

$(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}})$ and S are not independently distributed by the contrapositive of Rao's necessary condition. ‖

In view of Theorem 5.1, it follows that T_J^2 is not a form of Hotelling's T^2. Jackson's 75 base period observations could have been used to analyze the control of future observations as follows. Let $\bar{\underset{\sim}{X}}_{75}$ and S_{75} denote the sample mean vector and sample covariance matrix, respectively, based on the original 75 observations. Then for successive observations h, h = 1,2,..., the statistic

$$T^2 = (\underset{\sim}{X}_h-\bar{\underset{\sim}{X}}_{75})'S_{75}^{-1}(\underset{\sim}{X}_h-\bar{\underset{\sim}{X}}_{75})$$

is distributed as $\{(76)(74)p/(75)(73)\}F(p,75-p)$. This follows from Theorem 5.7 with n = 1. Even if Jackson had intended to use T_J^2 for the control of future observations, he apparently should have used the control limit $(5624p/5475)F(p,75-p,\alpha)$ and not $(74p/73)F(p,75-p,\alpha)$. For $p = 2$ and $\alpha = .05$, the suggested control limit is 6.4098 while Jackson's control limit is 6.3255. If Jackson assumed that the sample of size 75 was large enough to justify the replacement of $\bar{\underset{\sim}{X}}_{75}$ and S_{75} by $\underset{\sim}{\mu}$ and Σ, respectively, then the test statistic χ^2 is valid for testing the control of the original 75 observations and any future observations, where, for h = 1,2,...,75,...,

$$\chi^2 = (\underset{\sim}{X}_h-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X}_h-\underset{\sim}{\mu}).$$

The control limit now becomes $\chi^2(p,\alpha)$, which equals 5.99 for $p = 2$ and

$\alpha = .05$. Thus, if $\bar{X}_{75}$ and S_{75} are "close" to μ and Σ, respectively, then Jackson has used a good, heuristic procedure. Furthermore, Jackson's work is not to be slighted in view of his attempts to acquaint readers with Hotelling's T^2 and his utilization of principal components to reduce the dimensionality.

Ghare and Torgersen (23) have also apparently used the wrong test statistic. They wished to analyze 30 subgroups of 5 observations each for lack of control. Based on these 150 observations, they computed values for $\bar{\bar{X}}$ and S^{-1}. Evidently, they were unaware of any form of Hotelling's T^2 statistic since they used S as the population variance-covariance matrix and then used the following Chi-square test statistic:

$$\chi^2_{G,T} = n(\bar{X}_h - \bar{\bar{X}})' S^{-1} (\bar{X}_h - \bar{\bar{X}}),$$

for $h = 1,2,\ldots,30$. Obviously, $\chi^2_{G,T}$ is not distributed as a χ^2 random variable. It is not clear from their paper if S was computed using the formula $(1/149) \sum_{h=1}^{150} (X_h - \bar{\bar{X}})(X_h - \bar{\bar{X}})'$ or whether S was found by using the concept of a pooled estimate, denoted S_p. If S_p was used, then $n(\bar{X}_h - \bar{\bar{X}})' S_p^{-1} (\bar{X}_h - \bar{\bar{X}})$ is a form of Hotelling's T^2, and, for $\alpha = .005$, the control limit should be 10.80 and not 10.597 as they have stated. If the pooled estimate is not used, then $(\bar{X}_h - \bar{\bar{X}})$ and S are not independent and $n(\bar{X}_h - \bar{\bar{X}})' S^{-1} (\bar{X}_h - \bar{\bar{X}})$ is not a form of Hotelling's T^2. Also note that Ghare and Torgersen use X for $\bar{X}$ and V for Σ/n. However, they have apparently used for the first time the multivariate analogue of k rational subgroups.

More recently, Montgomery and Wadsworth (54) have also used a form of Hotelling's T^2. For ten rational subgroups with ten observations each and two quality characteristics, they use the test statistic

$$T^2_{M,W} = n(\bar{\underset{\sim}{X}}_h - \bar{\bar{\underset{\sim}{X}}})' S_p^{-1} (\bar{\underset{\sim}{X}}_h - \bar{\bar{\underset{\sim}{X}}}),$$

for $h = 1,2,\ldots,10$, with an upper control limit of $[p(n-1)/(n-p)]F(p,n-p,\alpha)$. For $k = 10$, $n = 10$, $p = 2$, and $\alpha = .01$, their control limit is 19.46, whereas it apparently should be $(1620/890)F(2,89,.01) = 8.82$. This corresponds more closely to $\chi^2(2,.01)$, which equals 9.21.

The first use of the T^2 random variable in a control chart setting was by Hotelling (36,37) in the testing of bombsights. However, his 1951 paper is primarily known for the development of the T_0^2 statistic, explained below. Let $\underset{\sim}{X}_{11}, \underset{\sim}{X}_{21}, \ldots, \underset{\sim}{X}_{m1}$ denote the elements of an initial random sample where $\bar{\underset{\sim}{X}}_m = (1/m) \sum_{h=1}^{m} \underset{\sim}{X}_{h1}$ and $S_m = (m-1)^{-1} \sum_{h=1}^{m} (\underset{\sim}{X}_{h1} - \bar{\underset{\sim}{X}}_m)(\underset{\sim}{X}_{h1} - \bar{\underset{\sim}{X}}_m)'$. Then a measure of the quality of a single bomb in a subsequent period is T_B^2, where

$$T_B^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \ell_{ij} X_{iB} X_{jB},$$

and ℓ_{ij} denotes the inverse of the sample variance-covariance matrix S_m. Obviously, T_B^2 is a form of Hotelling's T^2 mentioned in Theorem 5.1 with $p = 2$ and $\nu = m - 1$. An overall measure of the quality of a flight, sight, or lot is obtained by summing T_B^2 over the number of bombs involved for that characteristic. These are denoted by T_{OF}^2, T_{OS}^2, T_{OL}^2, respectively. These overall measures can be treated more generally.

Let $\underset{\sim}{X}_{12}, \underset{\sim}{X}_{22}, \ldots, \underset{\sim}{X}_{n2}$ denote the elements of a subsequent independent random sample where $\bar{\underset{\sim}{X}}_n = (1/n) \sum_{h=1}^{n} \underset{\sim}{X}_{h2}$. Then

$$T_0^2 = \sum_{h=1}^{n} T_{B_h}^2 = \sum_{h=1}^{n} \sum_{i=1}^{2} \sum_{j=1}^{2} \ell_{ij} X_{iB_h} X_{jB_h},$$

where T_0^2 does not have the Hotelling's T^2 distribution presented in Theorem 5.1. T_0^2 can be partitioned into T_M^2 and T_D^2, where

$$T_M^2 = n \sum_{i=1}^{2} \sum_{j=1}^{2} \ell_{ij} \bar{X}_i \bar{X}_j = n \bar{\underset{\sim}{X}}_n' S_m^{-1} \bar{\underset{\sim}{X}}_n$$

and

$$T_D^2 = \sum_{h=1}^{n} \sum_{i=1}^{2} \sum_{j=1}^{2} \ell_{ij} (X_{iB_h} - \bar{X}_i)(X_{jB_h} - \bar{X}_j) = \sum_{h=1}^{n} (\underset{\sim}{X}_{B_h} - \bar{\underset{\sim}{X}}_n)' S_m^{-1} (\underset{\sim}{X}_{B_h} - \bar{\underset{\sim}{X}}_n).$$

Now T_M^2 is a form of Hotelling's T^2 as in Theorem 5.1 with $p = 2$ and $v = m - 1$. However, T_D^2 is not of this form. Let S' denote the matrix with elements $s_{ij}' = (1/n) \sum_{h=1}^{n} X_{iB_h} X_{jB_h}$. Then $T_0^2 = n \sum_{i=1}^{2} \sum_{j=1}^{2} \ell_{ij} s_{ij}'$. Let S" denote the matrix with elements $s_{ij}'' = (n-1)^{-1} \sum_{h=1}^{n} (X_{iB_h} - \bar{X}_i)(X_{jB_h} - \bar{X}_j)$. Then $T_D^2 = (n-1) \sum_{i=1}^{2} \sum_{j=1}^{2} \ell_{ij} s_{ij}''$. Thus, both T_0^2 and T_D^2 are of the same form. To generalize the preceding cases, let S^* denote the matrix whose elements have q degrees of freedom. Then the statistic

$$T_0^2 = q \sum_{i=1}^{2} \sum_{j=1}^{2} \ell_{ij} s_{ij}^*$$

is known as Hotelling's Generalized T_0^2. The distribution of T_0^2 for

$p \geq 3$ has been studied by Constantine (12). T_0^2 can also be written as $q\ \mathrm{tr}(S^* S_m^{-1})$.

Hicks (30) has also used Hotelling's T^2 distribution in industrial applications, but control charts were not used. His primary use was to test the equality of two population means.

The only textbook presentation of the use of Hotelling's T^2 in a multivariate control chart setting has been by Johnson and Leone (46). For p variates, Johnson and Leone consider the control of subsequent individual observations by the statistic

$$T_{J,L}^2 = \left(\underset{\sim}{X}_i - E(\underset{\sim}{X})\right)' S^{-1} \left(\underset{\sim}{X}_i - E(\underset{\sim}{X})\right),$$

where $T_{J,L}^2$ is distributed as $\left[p\nu/(\nu-p+1)\right] F(p, \nu-p+1)$ with ν being the number of degrees of freedom in estimating the sample variances and covariances. They emphasize that S is based on past data. They also consider the case of replacing $\underset{\sim}{X}_i$ by $\bar{\underset{\sim}{X}}$. These cases correspond to some of the uses of Theorem 5.4.

This chapter commenced with a review of the empirical control charts for the mean for one quality characteristic and demonstrated how this could also be viewed as the familiar significance test for the equality of k means. For more than one quality characteristic, a test statistic was suggested for the case of rational subgroups and its distribution was developed. Using this test statistic, the marginal power for a subgroup was developed subject to certain conditions. The use of simultaneous techniques to determine the components responsible for rejection was also presented. For the concept of rational subgroups and for general p, these ideas do not seem to have previously appeared in the literature. The use of Hotelling's T^2 for other circumstances was also presented; for all of these cases, Σ was unknown, and it was desired to determine the control of the process relative to either some $\underset{\sim}{\mu}_0$ or some $\bar{\underset{\sim}{x}}_m$ determined from previous data. Some of these concepts also do not seem to have appeared previously.

CHAPTER VI

THE POWERS OF THE χ^2 AND T^2 TESTS

In testing the H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$ based on a random sample of size n from a normal population where σ is known, the power of the test is given by

$$\pi(\delta) = \Phi(\delta - z_{\alpha/2}) + \Phi(-\delta - z_{\alpha/2}), \tag{39}$$

where $\delta = (\mu_1 - \mu_0)\sqrt{n}/\sigma$ and $E(X) = \mu_1$. This agrees with the results presented in Chapter IV for p in general, where

$$\pi(\lambda) = P\left[n(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0) > \chi^2(p, \alpha) \,|\, \underset{\sim}{\mu}\right] \tag{40}$$

and $\lambda = n(\underset{\sim}{\mu} - \underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu} - \underset{\sim}{\mu}_0)$. Note that $\delta^2 = \lambda$. Because of the relationship between the noncentral χ^2 and noncentral F random variables, this power can be obtained from the Pearson and Hartley charts by letting $\nu_2 = \infty$. For the above-mentioned univariate test of hypothesis when σ is not known,

$$\pi(\delta) = P(\mathrm{mod}\,T'_{n-1} > t_{\alpha/2, n-1}), \tag{41}$$

where T'_{n-1} denotes the noncentral T random variable with n - 1 degrees of freedom. By definition,

$$T'_\nu = \frac{Z + \delta}{\sqrt{\chi^2(\nu)/\nu}} ,$$

where δ is some constant. Resnikoff and Lieberman (59) have prepared tables of its distribution. For the univariate test of hypothesis, $T'_{n-1} = T_{n-1} + (\delta\sigma/S)$, where $\sigma/S = \sqrt{(n-1)/\chi^2(n-1)}$. The power of this univariate test can be obtained directly from the Pearson and Hartley charts by letting $\nu_1 = 1$, $\nu_2 = n-1$, and $\phi = \text{mod}(\mu_1 - \mu_0)(n/2\sigma^2)^{1/2}$. This agrees with the results presented in Chapter V for general p where

$$\pi(\lambda) = P\left(n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'S^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0) > \left(p(n-1)/(n-p)\right)F(p,n-p,\alpha) \,|\, \underset{\sim}{\mu}\right) , \qquad (42)$$

and $\lambda = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$ with $\phi = \sqrt{\lambda/(p+1)}$.

Examination of the Pearson and Hartley charts reveals that for p in general the power is a monotonically increasing function of ϕ and thus of λ or δ^2. For p = 1, as $\text{mod}(\mu_1 - \mu_0)$ increases (decreases), δ^2 increases (decreases), and the power increases (decreases). Also as σ increases (decreases), δ^2 approaches zero (infinity), and the power approaches α (one). In the remainder of this chapter, interest is centered on the investigation of the power for general p, but particularly for p = 2. Since the noncentrality parameters for Equations (40) and (42) are identical, both of these cases will be considered simultaneously where Σ^{-1} in the latter case may be thought of as a nuisance parameter.

1. Properties of the Power Function

For $p = 2$, the noncentrality parameter $\lambda = n(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\underset{\sim}{\mu}-\underset{\sim}{\mu}_0)$ becomes

$$\lambda = n(1-\rho^2)^{-1}(a_1^2\sigma_1^{-2}-2\rho a_1 a_2\sigma_1^{-1}\sigma_2^{-1}+a_2^2\sigma_2^{-2}), \tag{43}$$

where $a_h = \mu_h - \mu_{h_0}$ and $E(\underset{\sim}{X}) = [\mu_1,\mu_2]'$. Note that for $i = 1,2, -\infty<a_i<\infty$, $\sigma_i>0$, and $-1<\rho<1$. The degenerate cases of $\sigma_i = 0$ or $\rho = 1$ will not be considered.

Equation (43) will first be considered as a function of a_1 and a_2, denoted g. That is, g: $\{(a_1,a_2): -\infty<a_i<\infty\} \to \{t: t\geq 0\}$, where $g(a_1, a_2) = n\underset{\sim}{a}'\Sigma^{-1}\underset{\sim}{a}$. Since $g(a_1,a_2)$ is a positive definite quadratic form in $\underset{\sim}{a}$, then g is a strictly convex function over all of E^2. Thus, the maximum of g is never attained, which implies that the power increases as $\|\underset{\sim}{a}-\underset{\sim}{0}\| = \|\underset{\sim}{\mu}-\underset{\sim}{\mu}_0\|$ increases. The minimum of g occurs when $\underset{\sim}{a} = \underset{\sim}{0}$ or $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$, in which case the power equals α, the probability of Type I error. These results agree with the univariate case. Quite often, the decision maker is interested in the effect of only one a_i. To indicate that one a_i is held constant, g will be denoted by $\hat{g}$. For constant a_2, $d\hat{g}/da_1 = n(1-\rho^2)^{-1}(-2\rho\sigma_1^{-1}\sigma_2^{-1}a_2+2a_1\sigma_1^{-2})$, and $d^2\hat{g}/da_2^2 = n(1-\rho^2)^{-1}(2\sigma_1^{-2})$, which indicate that the minimum power occurs at $a_1^0 = \rho\sigma_2^{-1}\sigma_1 a_2$. Furthermore,

$$\hat{g}(a_1^0,a_2) = na_2^2/\sigma_2^2 = \delta^2,$$

which is the square of the noncentrality parameter for a univariate test of hypothesis on the second component. For constant a_1, a similar minimum occurs. Morrison (55) obtains a similar result, assuming $\sigma_1 = 1$, $\sigma_2 = 1$. He also points out that "negative values of ρ lead to higher power probabilities for the same $\mu_i - \mu_{i_0}$ of like signs."

Equation (43) will now be considered as a function of σ_1 and σ_2, denoted f. That is, f: $\{(\sigma_1,\sigma_2)\colon \sigma_i>0\} \rightarrow \{t\colon t>0\}$, where the degenerate case $a_1 = a_2 = 0$ will not be considered. If $\underset{\sim}{a} = \underset{\sim}{0}$, then the range includes the zero point. Now the interpretation is that a_1 and a_2 are fixed. Setting the partials of f equal to zero implies that

$$\sigma_1 = a_1\rho^{-1}a_2^{-1}\sigma_2, \quad \sigma_2 = a_2\rho^{-1}a_1^{-1}\sigma_1.$$

Thus, there are no interior points at which the partials vanish.

Since

$$f(\sigma_1,\sigma_2) = n(1-\rho^2)^{-1}(a_1\sigma_1^{-1}-a_2\sigma_2^{-1})^2 + 2(1-\rho)a_1a_2\sigma_1^{-1}\sigma_2^{-1}, \qquad (44)$$

it follows that $f(\sigma_1,\sigma_2) \rightarrow \infty$ as $(\sigma_1,\sigma_2) \rightarrow (0,0)$. Furthermore, $f(\sigma_1,\sigma_2) \rightarrow 0$ as $(\sigma_1,\sigma_2) \rightarrow (\infty,\infty)$. These results agree with the univariate case. It should be noted that f is neither strictly convex nor strictly concave. The properties of f will now be investigated by fixing either σ_1 or σ_2 and considering different cases. To indicate that either σ_1 or σ_2 is fixed, f will be denoted by $\hat{f}$.

The first case to be considered is when a_1 and a_2 have the same signs. Suppose $\rho = 0$. Then $f(\sigma_1,\sigma_2) = n(a_1^2\sigma_1^{-2}+a_2^2\sigma_2^{-2})$. Fix σ_j. For

$i \neq j$, as σ_i increases, $\hat{f}$ decreases to na_j^2/σ_j^2, which equals δ^2, the noncentrality parameter for a univariate test of hypothesis on the *jth* component. As σ_i decreases, $\hat{f}$ increases, and the power increases. Suppose $\rho < 0$, and $f(\sigma_1,\sigma_2)$ is given by Equation (43). As Morrison has pointed out, the power is automatically increased since all the terms are positive. Fix σ_j. For $i \neq j$, as σ_i increases (decreases), $\hat{f}$ decreases (increases), and the power decreases (increases). This also applies if σ_1 and σ_2 simultaneously increase (decrease). Finally, suppose $\rho > 0$. Again $f(\sigma_1,\sigma_2)$ is given by Equation (43). Fix σ_2. Then $d\hat{f}/d\sigma_1 = n(1-\rho^2)^{-1}(-2a_1^2\sigma_1^{-3}+2\rho a_1a_2\sigma_1^{-2}\sigma_2^{-1})$ and $d^2\hat{f}/d\sigma_1^2 = n(1-\rho^2)^{-1}(6a_1^2\sigma_1^{-4}-4\rho a_1a_2\sigma_1^{-3}\sigma_2^{-1})$. Let $\hat{f}'$ and $\hat{f}''$ denote $d\hat{f}/d\sigma_1$ and $d^2\hat{f}/d\sigma_1^2$, respectively. Setting $\hat{f}'(\sigma_1,\sigma_2) = 0$ implies that $\sigma_1^0 = a_1\sigma_2\rho^{-1}a_2^{-1}$, where σ_1^0 denotes a critical point. Since $\hat{f}''(\sigma_1^0,\sigma_2)>0$, $\hat{f}(\sigma_1^0,\sigma_2)$ is a minimum value of $\hat{f}$. Furthermore, let σ_1^* denote the point of inflection of $\hat{f}$; then $\sigma_1^* = (3/2)a_1\sigma_2\rho^{-1}a_2^{-1}$. This follows since $\hat{f}''(\sigma_1,\sigma_2)>0$ for $\sigma_1<\sigma_1^*$, $\hat{f}''(\sigma_1^*,\sigma_2) = 0$, and $\hat{f}''(\sigma_1,\sigma_2)<0$ for $\sigma_1>\sigma_1^*$. Also note that

$$\hat{f}(\sigma_1^0,\sigma_2) = na_2^2\sigma_2^{-2} = \delta^2 \tag{45}$$

$$\hat{f}(\sigma_1^*,\sigma_2) = na_2^2\sigma_2^{-2}\{(9-8\rho^2)/(9-9\rho^2)\}$$

$$\hat{f}(\infty,\sigma_2) = na_2^2\sigma_2^{-2}(1-\rho^2)^{-1}$$

Furthermore, σ_1^0 is a global minimum since there are no local maxima. These properties are shown in Figure 6 for two different values of ρ.

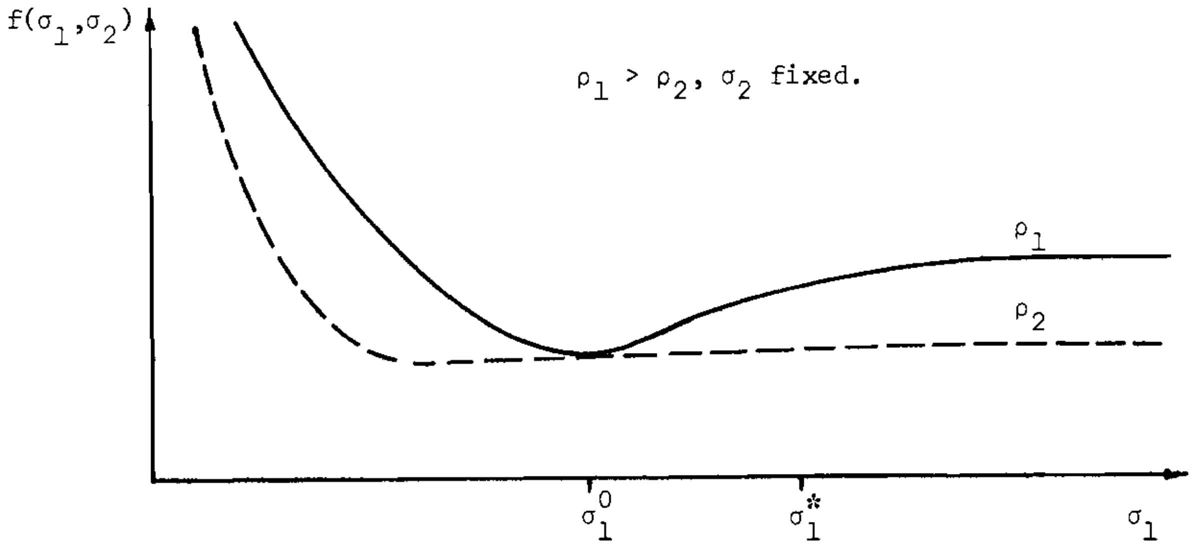

Figure 6. An Illustration of the Power Paradox

Similar results are obtained for σ_2 by fixing σ_1. In this case, $\sigma_2^0 = a_2\sigma_1\rho^{-1}a_1^{-1}$. Thus f has the following properties when a_1 and a_2 have the same sign and σ_2 is fixed. For $\rho = 0$, $\hat{f}(\sigma_1,\sigma_2) \to \delta^2$ as $\sigma_1 \to \infty$. For $\rho < 0$, $\hat{f}(\sigma_1,\sigma_2) \to (1-\rho^2)^{-1}\delta^2$ as $\sigma_1 \to \infty$. For $\rho > 0$, $\hat{f}(\sigma_1^0,\sigma_2) = \delta^2$, which is less than $\hat{f}(\infty,\sigma_2) = (1-\rho^2)^{-1}\delta^2$.

This final case leads to the following paradox. For fixed σ_2 and increasing σ_1, the noncentrality parameter and the power decrease for $\sigma_1<\sigma_1^0$; the power is a minimum when $\sigma_1 = \sigma_1^0$; and, for $\sigma_1>\sigma^0$, the power asymptotically increases as σ_1 increases. Thus, for fixed σ_2, the noncentrality parameter is not a monotonically decreasing function of σ_1 as it was in the univariate case. Furthermore, at σ_1^0, the value of the noncentrality parameter is given by the univariate noncentrality parameter for σ_2, as shown in Equation (45). Two resolutions of this paradox will be offered in the next section. This result is also extremely important to the decision maker when because of costs or other constraints only one σ_i can be adjusted. A general policy would be to always decrease σ_i when $\sigma_i < \sigma_i^0$. However, it is also permissible to increase σ_i when $\sigma_i > \sigma_i^0$. But, always avoid setting $\sigma_i = \sigma_i^0$.

The next case to be considered is when a_1 and a_2 have different signs. For $\rho = 0$, the same results are obtained as before. For $\rho > 0$, the same results are obtained as before when a_1 and a_2 had the same sign and $\rho < 0$. Finally, for $\rho < 0$, the same results are obtained as before when $\rho > 0$.

2. Resolutions of the Paradox

The first interpretation of the paradox requires the concept of Fisherian information. Since this resolution of the paradox is not directly related to the major theme of quality control, the necessary background will be presented in an informal manner. The definitions and other concepts are based on those presented by Hogg and Craig (34), Lindgren (52), and Silvey (64).

To provide an introduction, the first case to be considered is when the parameter θ is one-dimensional and the random variable X is also one-dimensional. Let $\{f_X(x;\theta)\colon \theta\epsilon\Omega\}$ denote a family of probability density functions, where $f_X(x;\theta)$ is of known functional form but the parameter θ is unknown. For a fixed positive integer n, let $\underset{\sim}{X}' = [X_1, X_2, \ldots, X_n]$ denote a random sample of size n from a distribution that is one member of the family $\{f_X(x;\theta)\colon \theta\epsilon\Omega\}$. The basic problem in point estimation is that of defining a statistic $U = h(\underset{\sim}{X}')$ so that $u = h(\underset{\sim}{x}')$

will be a "good" point estimate of θ. Let $U_1 = h_1(\underset{\sim}{X}')$ be another estimator of θ. The criterion of efficiency is sometimes used to compare these two estimators, where the estimator U is said to be more efficient than U_1 if $E\left((U-\theta)^2\right) \le E\left((U_1-\theta)^2\right)$, and the relative efficiency of U_1 with respect to U, denoted $e(U_1,U)$, is the ratio of these quadratic loss functions. As Lindgren states, "An absolute measure of efficiency of an estimate would require that its mean square deviation from the parameter being estimated be compared with a lower bound or absolute minimum of such mean square deviations, if one that is not zero exists." The Cramér-Rao inequality, also known as the Frechét inequality, provides such a lower bound. Let $X_1,X_2,\ldots,X_n$ denote a random sample of size n from a distribution that is one member of the family $\{f_X(x;\theta)\colon \theta\epsilon\Omega\}$, and let $U = h(\underset{\sim}{X}')$ be considered as an estimator for the parameter θ, then by the Cramér-Rao inequality,

$$V(U) \ge \left(1+b_U'(\theta)\right)^2/V(V),$$

where $V = \partial\left(\ell n f_{\underset{\sim}{X}'}(\underset{\sim}{X}';\theta)\right)/\partial\theta$, $b_U(\theta)$ is the bias in U, and $b_U'(\theta)$ is the derivative of $b_U(\theta)$ with respect to θ. Certain regularity conditions must be satisfied. Since $E(V) = 0$, $V(V) = E(V^2)$. The quantity $E(V^2)$ was called by Fisher the amount of information about θ contained in the sample. It is often denoted by $I(\theta)$. If U is an unbiased estimator of θ, then the inequality becomes $V(U)\ge 1/I(\theta)$. As Silvey states, "The more information about θ provided on average by the sample, the smaller we might expect the variance of a 'good' estimator to be." $I(\theta)$ has two very useful properties. The first is that $I(\theta) = E(V^2) = -E(\partial V/\partial\theta)$. Let $\underset{\sim}{X}_1'$ and $\underset{\sim}{X}_2'$ be two independent random samples from a distribution that is one member of the family $\{f_X(x;\theta): \theta\epsilon\Omega\}$. Then the second property is that $I_{\underset{\sim}{X}_1',\underset{\sim}{X}_2'}(\theta) = I_{\underset{\sim}{X}_1'}(\theta) + I_{\underset{\sim}{X}_2'}(\theta)$. Thus, if $i(\theta)$ denotes the information about θ contained in a single observation, $I(\theta) = ni(\theta)$ based on a random sample of size n.

These concepts will now be extended to the case where $\underset{\sim}{\theta}$ is an (r×1) vector of parameters and the random variable $\underset{\sim}{X}$ is (p×1). For a fixed positive integer n, let $X = [\underset{\sim}{X}_1,\underset{\sim}{X}_2,\ldots,\underset{\sim}{X}_n]$ denote a random sample of size n from a distribution that is one member of the family $\{f_{\underset{\sim}{X}'}(\underset{\sim}{x}';\underset{\sim}{\theta}): \underset{\sim}{\theta}\epsilon\Omega\}$, where $f_{\underset{\sim}{X}'}(\underset{\sim}{x}';\underset{\sim}{\theta})$ satisfies certain regularity conditions. Let the (r×1) vector $\underset{\sim}{U} = [U_1,U_2,\ldots,U_r]'$ be an unbiased estimator of $\underset{\sim}{\theta} = [\theta_1,\theta_2,\ldots,\theta_r]'$, and let the (r×1) vector $\underset{\sim}{V}$ have components

$$V_h = \partial\left(\ell n f_{\underset{\sim}{X}_1',\ldots,\underset{\sim}{X}_n'}(\underset{\sim}{X}_1',\ldots,\underset{\sim}{X}_n';\underset{\sim}{\theta})\right)/\partial\theta_h,$$

for $h = 1,2,\ldots,r$. Furthermore, let I_θ be the (r×r) matrix where $I_\theta = [E(V_iV_j)]$, and $E(V_iV_j) = -E(\partial V_i/\partial\theta_j)$. If $C(\underset{\sim}{U},\underset{\sim}{U}')$ denotes the covariance matrix of $\underset{\sim}{U}$, then the generalized Cramér-Rao inequality states that $C(\underset{\sim}{U},\underset{\sim}{U}') - I_\theta^{-1}$ is positive semi-definite. As Silvey states, I_θ^{-1} is "in a sense a 'lower bound' for the variance matrix of an unbiased estimator of $\underset{\sim}{\theta}$." Thus, the matrix I_θ is a multivariate analogue of the information, where $C(\underset{\sim}{V},\underset{\sim}{V}') = I_\theta$. Silvey proves the generalized information inequality when $E(\underset{\sim}{U}) = \underset{\sim}{\theta}$. Apparently, this result can be generalized to consider the class of biased estimators. That is, for $h = 1,2,\ldots,r$, let $E(U_h) = \theta_h + b(\theta_h)$, where $b(\theta_h)$ is the bias in U_h when used as an estimator for θ_h. Also, let $b'(\theta_h)$ be the derivative of $b(\theta_h)$ with respect to θ_h, and let D be an $(r\times r)$ diagonal matrix where the diagonal elements are $1 + b'(\theta_h)$. The Cramér-Rao lower bound now states that the matrix $C(\underset{\sim}{U},\underset{\sim}{U}') - DI_\theta^{-1}D$ is positive semi-definite. Since the noncentrality parameter in Equation (43) pertained to sampling from a normal population, only this case will be considered.

Since Σ^{-1} is assumed to be constant, the $(p\times 1)$ vector $\underset{\sim}{\theta} = \underset{\sim}{\mu}$. To preserve a single notation, let L and $\ell n L$ be defined as in Equations (2) and (3) with the understanding that the likelihood interpretation does not apply here. In the proof of Theorem 3.1, it was shown that $\partial \ell n L/\partial \underset{\sim}{\mu} = n\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})$. Thus, $\partial^2 \ell n L/\partial \underset{\sim}{\mu}\partial \underset{\sim}{\mu}' = -n\Sigma^{-1}$, and the information matrix is $-E(-n\Sigma^{-1}) = n\Sigma^{-1}$, while $I_\theta^{-1} = \Sigma/n$ is a "lower bound" for the variance-covariance matrix of an unbiased estimator of $\underset{\sim}{\mu}$. In Theorem 3.2, it was shown that $\bar{\underset{\sim}{X}}$ is an unbiased estimator of $\underset{\sim}{\mu}$ and that $C(\bar{\underset{\sim}{X}},\bar{\underset{\sim}{X}}') = \Sigma/n$ so that the "lower bound" is attained in this case. For $p = 1$, it follows that $I_\theta = n\sigma^{-2}$, while, for $p = 2$,

$$I_\theta = \frac{n}{1-\rho^2}\begin{bmatrix}\sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2}\end{bmatrix}. \tag{46}$$

Let $L(\underset{\sim}{X})$ denote the probability law of the random variable $\underset{\sim}{X}$. Theorem 2.4 stated that if $L(\underset{\sim}{X}) = N(\underset{\sim}{\mu},\Sigma)$, then $L(\underset{\sim}{X}_1|\underset{\sim}{X}_2=\underset{\sim}{x}_2) = N\left(\underset{\sim}{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2), \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$. For $p = 2$ this becomes $L(X_1|X_2=x_2) = N\left(\mu_1+\rho(\sigma_1/\sigma_2)(x_2-\mu_2),\sigma_1^2(1-\rho^2)\right)$. If σ_1,σ_2, and ρ are constant and X_2 is fixed, then the information matrix for μ_1 and μ_2 based on a random sample of size n is given by

$$I_\theta = \frac{n}{1 - \rho^2} \begin{bmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \rho^2\sigma_2^{-2} \end{bmatrix} \tag{47}$$

To distinguish between the notations used in Equations (46) and (47), let $I(X_1,X_2)$ and $I(X_1|X_2)$ denote Equations (46) and (47), respectively. Furthermore, let $I(X_2)$ denote the information for μ_2 in a random sample of size n from the univariate normal density of X_2. It was shown that $I(X_2) = n\sigma_2^{-2}$. Now

$$I(X_1,X_2) = \frac{n}{1 - \rho^2} \begin{bmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \rho^2\sigma_2^{-2} \end{bmatrix} + n \begin{bmatrix} 0 & 0 \\ 0 & \sigma_2^{-2} \end{bmatrix}.$$

Thus, $I(X_1,X_2) = I(X_1|X_2) + I(X_2)$.

One final property of Fisher's information needs to be presented for the multiparameter case. If $\underset{\sim}{U}$ is an unbiased estimator of $\underset{\sim}{\theta}$, then $\underset{\sim}{c}'\underset{\sim}{U}$ is an unbiased estimator of $\underset{\sim}{c}'\underset{\sim}{\theta}$, where $\underset{\sim}{c}$ is a (r×1) vector of known constants. Since $C\left(\underset{\sim}{c}'\underset{\sim}{U},(\underset{\sim}{c}'\underset{\sim}{U})'\right) = \underset{\sim}{c}'C(\underset{\sim}{U},\underset{\sim}{U}')\underset{\sim}{c}$, it follows that $\underset{\sim}{c}'C(\underset{\sim}{U},\underset{\sim}{U}')\underset{\sim}{c} \geq \underset{\sim}{c}'I_\theta^{-1}\underset{\sim}{c}$. Silvey states that if we can find an unbiased estimator $\underset{\sim}{U}$ such that $C(\underset{\sim}{U},\underset{\sim}{U}') = I_\theta^{-1}$ this estimator has in this linear sense smaller dispersion than any other unbiased estimator of $\underset{\sim}{\theta}$. Thus, for the bivariate normal, $\underset{\sim}{a}'\underset{\sim}{\bar{X}}$ has smaller dispersion than any other linear unbiased estimator of $\underset{\sim}{\mu}$, and this dispersion is given by $\underset{\sim}{a}'I(X_1,X_2)\underset{\sim}{a}$ which equals

$$n(1-\rho^2)^{-1}(a_1^2\sigma_1^{-2}-2\rho a_1a_2\sigma_1^{-1}\sigma_2^{-1}+a_2^2\sigma_2^{-2}). \tag{48}$$

Since $I(X_1,X_2) = I(X_2) + I(X_1|X_2)$, it follows that $\underset{\sim}{a}'\left(I(X_2)+I(X_1|X_2)\right)\underset{\sim}{a} = \underset{\sim}{a}'I(X_1,X_2)\underset{\sim}{a}$, where

$$\underset{\sim}{a}'\left(I(X_2)+I(X_1|X_2)\right)\underset{\sim}{a} = n(1-\rho^2)^{-1}(a_1\sigma_1^{-1}-a_2\rho\sigma_2^{-1})^2 + na_2^2\sigma_2^{-2}, \tag{49}$$

and Equations (48) and (49) are identical. Since Equation (49) is the sum of two nonnegative terms and $na_2^2\sigma_2^{-2}$ is fixed, the information as presented in Equation (49) is minimized when $(a_1\sigma_1^{-1}-a_2\rho\sigma_2^{-1})^2 = 0$ or when $\sigma_1 = a_1\sigma_2\rho^{-1}a_2^{-1}$. For $i = 1,2$, let $a_i = \mu_i - \mu_{i_0}$. Then the total linear information given by Equation (48) is identical with the noncentrality parameter presented in Equation (43). Furthermore, for a_1 and a_2 of the same sign, $\rho > 0$, and fixed σ_2, that value of σ_1 which minimizes the noncentrality parameter is that same value of σ_1 which minimizes the information when X_2 is fixed. Thus, this is one interpretation for the minimization of the noncentrality parameter at σ_1^0.

In the noncentrality parameter for Equation (42), Σ^{-1} was considered as a nuisance parameter. Actually, since $\theta = [\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho]'$, the information matrix for this case is a (5×5) block diagonal matrix where the diagonal matrices are (2×2) and (3×3) and the off diagonal matrices are (2×3) and (3×2) zero matrices. The (2×2) diagonal matrix is identical with Equation (46). Thus, if one lets the (5×1) vector $\underset{\sim}{a} = [a_1, a_2, 0, 0, 0]'$, then the same results are obtained.

When the noncentrality parameter was considered as a function of a_1 for fixed a_2, it was found that $a_1^0 = \rho\sigma_2^{-1}\sigma_1 a_2$. This is the same value of a_1 that minimizes theinformation presented in Equation (49).

The second resolution of the paradox is in the framework of a test of hypothesis. Let $\underset{\sim}{Y}$ be a (p×1) vector where $\underset{\sim}{Y} \sim N(\underset{\sim}{\mu}_{\underset{\sim}{Y}}, \Sigma)$ and Σ is known. It is of interest to test

$$H_0\colon \underset{\sim}{\mu}_{\underset{\sim}{Y}} = \underset{\sim}{\mu}_0 \quad \text{vs.} \quad H_1\colon \underset{\sim}{\mu}_{\underset{\sim}{Y}} \neq \underset{\sim}{\mu}_0.$$

Let $\underset{\sim}{X} = R^{-1}(\underset{\sim}{Y} - \underset{\sim}{\mu}_0)$, where $\Sigma = RR'$. Then $\underset{\sim}{X}$ is observable and $\underset{\sim}{X} \sim N\big(R^{-1}(\underset{\sim}{\mu}_{\underset{\sim}{Y}} - \underset{\sim}{\mu}_0), I\big)$. For the transformed problem,

$$H_0\colon \underset{\sim}{\mu}_{\underset{\sim}{X}} = \underset{\sim}{0} \quad \text{vs.} \quad H_1\colon \underset{\sim}{\mu}_{\underset{\sim}{X}} \neq \underset{\sim}{0}. \tag{50}$$

Thus, in this second interpretation of the paradox, it suffices to consider hypotheses of the form stated in Equation (50). Furthermore, only the case of p = 2 will be considered. Also note that $L(\underset{\sim}{X}) = L(X_1, X_2) = L(X_1|X_2{=}x_2)L(X_2)$, in view of the definition of a conditional probability density.

Under H_0 given in Equation (50), $L(\underset{\sim}{X}) = L(X_1|X_2{=}x_2)L(X_2) = N\big(\rho(\sigma_1/\sigma_2)x_2, \sigma_1^2(1-\rho^2)\big)N(0, \sigma_2^2)$. Under H_1, $L(X_1|X_2{=}x_2)L(X_2) = N\big(\mu_1 + \rho(\sigma_1/\sigma_2)(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\big)N(\mu_2, \sigma_2^2)$. $L(X_1|X_2{=}x_2)$ under H_1 is identical to $L(X_1|X_2{=}x_2)$ under H_0 if $\mu_1 - \rho(\sigma_1/\sigma_2)\mu_2 = 0$, or if $\sigma_1 = \mu_1\sigma_2\rho^{-1}\mu_2^{-1}$, which is the same value of σ_1 that minimizes the noncentrality parameter. Thus, for testing H_0, no additional information is obtained by looking at X_1.

This chapter has introduced a paradox and partial resolutions associated with the tests of significance presented in Chapters IV and V. After an extensive literature search, it seems that neither the paradox nor its resolutions have appeared previously in the literature.

CHAPTER VII

CONTROL CHARTS FOR DISPERSION

In both Chapters IV and V, it was assumed that the process dispersion remained constant. This assumption must be validated in practice. The dispersion of the process is usually controlled by sigma charts or range charts. Primary emphasis in this chapter will be devoted to both the theoretical and empirical sigma charts and their analogues, with a brief mention of the range chart. After the univariate dispersion charts have been presented, their multivariate counterparts will be presented.

1. Univariate Dispersion Control Charts

Theoretical Charts

Theoretical control charts will be considered first. It is desired that the process remain at the nominal value σ_0. This value could have been derived from past data or selected by management to attain certain objectives. To determine whether the process dispersion is in control at a given time, a random sample of size n is obtained and a realization of some statistic is determined from this sample data.

The first chart to be considered is the S^2-chart. Let $X_1,\ldots,X_n$ be a random sample from X, where $X \sim N(\mu,\sigma_0^2)$. Define $S^2 = (n-1)^{-1} \sum_{h=1}^{n} (X_h-\bar{X})^2$. Then $(n-1)S^2/\sigma_0^2 \sim \chi^2(n-1)$, and

$$P\{\sigma_0^2\chi^2(n-1,1-\alpha/2)/(n-1)\le S^2\le\sigma_0^2\chi^2(n-1,\alpha/2)/(n-1)\} = 1 - \alpha.$$

Thus, a theoretical control chart for the sample variance would have an upper control limit given by $\sigma_0^2\chi^2(n-1,\alpha/2)/(n-1)$ and a lower control limit given by $\sigma_0^2\chi^2(n-1,1-\alpha/2)/(n-1)$. It is customary to use only an upper control limit replacing $\chi^2(n-1,\alpha/2)$ by $\chi^2(n-1,\alpha)$. Also, a control chart for the sample standard deviation would have its control limits given by the square roots of the control limits for the S^2-chart.

This could also be viewed as a hypothesis testing problem. The decision maker would be interested in testing H_0: $\sigma^2 = \sigma_0^2$ vs. H_1: $\sigma^2 \neq \sigma_0^2$. Specifically, Ω and ω are those subsets of the Euclidean plane such that

$$\Omega = \{(\mu,\sigma^2): -\infty<\mu<\infty,\ \sigma^2>0\}$$

and

$$\omega = \{(\mu,\sigma^2): -\infty<\mu<\infty,\ \sigma^2=\sigma_0^2\}.$$

By obtaining $\partial \ln L(\omega)/\partial\mu$, equating this to zero, and setting $\sigma^2 = \sigma_0^2$, one obtains $\hat{\mu}_\omega = \bar{x}$, $\hat{\sigma}_\omega^2 = \sigma_0^2$, and

$$L(\hat{\omega}) = (\sigma_0^2)^{-n/2}(2\pi)^{-n/2}\exp\{-(1/2\sigma_0^2)\sum_{h=1}^{n}(x_h-\bar{x})^2\}.$$

Similarly, one obtains $\hat{\mu}_\Omega = \bar{x}$, $\hat{\sigma}_\Omega^2 = (1/n)\sum_{h=1}^{n}(x_h-\bar{x})^2$, and

$$L(\hat{\Omega}) = \left(n^{-1} \sum_{h=1}^{n} (x_h - \bar{x})^2\right)^{-n/2} (2\pi)^{-n/2} \exp(-n/2).$$

Thus, for $\underset{\sim}{x}' = [x_1, x_2, \ldots, x_n]$, $\lambda(\underset{\sim}{x}') = (u/n)^{n/2} \exp\left[-(n/2)\left((u/n)-1\right)\right]$, where $u = \sigma_0^{-2} \sum_{h=1}^{n} (x_h - \bar{x})^2$, and the form of the critical region is given by $w = \{\underset{\sim}{x}' \colon \lambda(\underset{\sim}{x}') < k\}$. Since U is distributed as $\chi^2(n-1)$, the distribution of $\lambda(\underset{\sim}{X}')$ could be found by transforming u to $(u/n)^{n/2} \exp\left[-(n/2)\left((u/n)-1\right)\right]$. Instead, the critical region will be expressed in a different form. Let $y = \left(\lambda(\underset{\sim}{x}')\right)^{2/n} e^{-1}$. Then $y = (u/n)\exp\left(-(u/n)\right)$, and the form of the critical region is given by $w = \{\underset{\sim}{x}' \colon y < k_1\}$, where $k_1 = k^{2/n} e^{-1}$. By examining the derivative of y with respect to u, it can be seen that y is not a monotonic function of u. In fact, y increases for $u < n$, reaches a maximum at $u = n$, and decreases for $u > n$. Therefore, an equivalent critical region in terms of u is given by

$$w = \{\underset{\sim}{x}' \colon 0 < u < k_2\} \cup \{\underset{\sim}{x}' \colon k_3 < u < \infty\},$$

where k_2 and k_3 are chosen so that the total probability of type I error is α. Since $U \sim \chi^2(n-1)$, then

$$w = \{\underset{\sim}{x}' \colon 0 < u < \chi^2(n-1, 1-\alpha/2)\} \cup \{\underset{\sim}{x}' \colon \chi^2(n-1, \alpha/2) < u < \infty\}.$$

Note that the rejection regions are equivalent to the regions above the UCL and below the LCL for the S^2-chart since $(n-1)S^2/\sigma_0^2 = U$. Given that $\sigma^2 = \sigma_1^2 \neq \sigma_0^2$, the power of the test is given by

$$\pi(\lambda) = 1 - P\left(\lambda^{-2}\chi^2(n-1,1-\alpha/2) \le \chi^2(n-1) \le \lambda^{-2}\chi^2(n-1,\alpha/2)\right)$$

where $\lambda = \sigma_1/\sigma_0$. The determination of the power is facilitated by the charts in Bowker and Lieberman (8). The S^2-chart or the equivalent test of significance can be considered an exact procedure.

Other theoretical dispersion charts do not make full use of the distribution of the sample statistic, but only use the first two population moments. Based on a random sample of size n, let $S = \sqrt{S^2}$. For r a member of the positive integers and since $(n-1)S^2/\sigma_0^2 \sim \chi^2(n-1)$, it follows that

$$E(S^r) = \sigma_0^r (n-1)^{-r/2} E\left[\left((n-1)S^2/\sigma_0^2\right)^{r/2}\right]$$

$$= \sigma_0^r \left(2/(n-1)\right)^{r/2} \Gamma\left((n+r-1)/2\right)/\Gamma\left((n-1)/2\right)$$

An immediate consequence is that

$$E(S) = \sigma_0 \left(2/(n-1)\right)^{1/2} \Gamma(n/2)/\Gamma\left((n-1)/2\right) = \sigma_0 c_2', \quad \text{and} \quad E(S^2) = \sigma_0^2.$$

Thus,

$$V(S) = \sigma_0^2 (1 - c_2'^2).$$

These moments could have also been found by first finding the probability law of S, which is proportional to a χ-random variable. Since most of the probability distribution of S is contained in the interval $E(S) \pm 3\sqrt{V(S)}$, it seems reasonable that a chart for S should have limits given by

$$\mathrm{UCL} = \sigma_0\left(c_2' + 3(1-c_2'^2)^{1/2}\right) \tag{51}$$

$$\mathrm{CL} = \sigma_0 c_2' \tag{52}$$

$$\mathrm{LCL} = \sigma_0\left(c_2' - 3(1-c_2'^2)^{1/2}\right). \tag{53}$$

It is customary to replace the lower control limit by 0 if $c_2' - 3(1-c_2'^2)^{1/2} < 0$, which occurs for $n < 6$. Since S is not normally distributed, these limits cannot be thought of as .9973 probability limits. The control chart quantities given by Equations (51) through (53) differ from the traditional limits since $S = \sqrt{S^2}$ was used in place of $V = \sqrt{V^2}$, where $V^2 = (1/n)\sum_{h=1}^{n}(X_h - \bar{X})^2$. Since

$$E(V^r) = \sigma_0^r (2/n)^{r/2}\Gamma\left((n+r-1)/2\right)/\Gamma\left((n-1)/2\right),$$

it immediately follows that

$$E(V) = \sigma_0(2/n)^{1/2}\Gamma(n/2)/\Gamma\left((n-1)/2\right) = \sigma_0 c_2$$

and

$$\mathcal{V}(V) = (\sigma_0^2/n)\left[(n-1)-nc_2^2\right].$$

Thus, the upper and lower control limits for V are given by

$$\sigma_0\left[c_2 \pm 3n^{-1/2}(n-1-nc_2^2)^{1/2}\right],$$

where, for large n, there is negligible difference between these limits and those given by Equations (51) and (53) since $c_2 = c_2'(1-n^{-1})^{1/2}$. The control chart for V is usually called the sigma chart.

The most popular of the dispersion charts is the range chart, especially for small sample sizes. Its popularity stems from its simplicity. However, it is inefficient since it does not use all of the elements of the sample. Let $X \sim N(\mu,\sigma_0^2)$, and let $X_1,X_2,\ldots,X_n$ be a random sample from X. Define $R = \max\{X_1,X_2,\ldots,X_n\} - \min\{X_1,X_2,\ldots,X_n\} = X_{(n)} - X_{(1)}$. Then, it can be shown that

$$f_R(r) = n(n-1)\int_{-\infty}^{\infty}\left[F_X(u+r)-F_X(u)\right]^{n-2}f_X(u)f_X(u+r)du,\ r > 0,$$

with $E(R) = \sigma_0 d_2$ and $\mathcal{V}(R) = d_3^2\sigma_0^2$. Tables of d_2 and d_3 are given in Bowker and Lieberman (8) for n = 2(1)25. Since the distribution of R does not depend on μ and the dependence on σ_0 is relatively simple, the distribution of $W = R/\sigma_0$ is usually considered where $f_W(w) = \sigma_0 f_R(\sigma_0 w)$. Since most of the distribution of R is contained in the interval $E(R) \pm 3\sqrt{\mathcal{V}(R)}$, one method of constructing a control chart for R would have upper and lower control limits given by $\sigma_0(d_2+3d_3)$ and $\sigma_0(d_2-3d_3)$, respectively, and a center line given by $\sigma_0 d_2$. The factors $D_1 = d_2 - 3d_3$ and $D_2 = d_2 + 3d_3$ are given in Bowker and Lieberman (8) for n = 2(1)25. A control chart for R could also be constructed using the percentage points of W. Let $w_{\alpha,n}$ and $w_{1-\alpha,n}$ denote the upper and lower α-percentage points of W, respectively. Since $P(w_{1-\alpha/2,n} \leq W \leq w_{\alpha/2,n}) = 1 - \alpha$, upper and lower control limits for R would be given by $\sigma_0 w_{\alpha/2,n}$ and $\sigma_0 w_{1-\alpha/2,n}$, respectively. Percentage points of W are given in Duncan (18) for n = 2(1)12 and α = .001, .005, .010, .025, and .050.

Empirical Charts

Assume that k rational subgroups of n observations each have been collected, where statistical control existed within each subgroup. Refer to Table 8.

Table 8. Data for p = 1

Sample Number	Individual Values	Subgroup Statistics	Population Parameters
1	$x_{11}, x_{12}, \ldots, x_{1n}$	$\bar{x}_1, s_1^2, s_1$	μ_1, σ_1^2
2	$x_{21}, x_{22}, \ldots, x_{2n}$	$\bar{x}_2, s_2^2, s_2$	μ_2, σ_2^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$x_{k1}, x_{k2}, \ldots, x_{kn}$	$\bar{x}_k, s_k^2, s_k$	μ_k, σ_k^2

Note that Table 8 differs from Table 4 in that now it must be determined whether $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma_0^2$, where σ_0^2 is unspecified. Most of the procedures to be presented are ad hoc procedures.

The first empirical dispersion control chart to be considered is the analogue of the theoretical S^2 chart. For $h = 1,2,\dots,k$, let $s_h^2 = (n-1)^{-1} \sum_{j=1}^{n} (x_{hj}-\bar{x}_h)^2$, where $\bar{x}_h = (1/n) \sum_{j=1}^{n} x_{hj}$. Then $(n-1)S_h^2/\sigma_h^2$ is distributed $\chi^2(n-1)$ for $h = 1,2,\dots,k$. If $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma_0^2$, then $(1/\sigma_0^2) \sum_{h=1}^{k} (n-1)S_h^2$ is distributed $\chi^2\big(k(n-1)\big)$ and $S_p^2 = (1/k) \sum_{h=1}^{k} S_h^2$ is an unbiased estimator of σ_0^2. When the within group sample sizes are unequal, $S_p^2 = \sum_{h=1}^{k} \nu_h S_h^2 / \sum_{h=1}^{k} \nu_h$, where $\nu_h = (n_h-1)$. Since theoretical limits were given by $\sigma_0^2\chi^2(n-1,1-\alpha/2)/(n-1)$ and $\sigma_0^2\chi^2(n-1,\alpha/2)/(n-1)$ and s_p^2 is an unbiased estimate of σ_0^2, reasonable control limits for S_h^2 would be given by $s_p^2\chi^2(n-1,1-\alpha/2)/(n-1)$ and $s_p^2\chi^2(n-1,\alpha/2)/(n-1)$. One advantage of these control limits is that they are easily computed. An obvious disadvantage is that s_p^2 replaces σ_0^2 regardless of the number of subgroups, and the distributional properties of S_p^2 are not utilized. However, it seems that S_p^2 is a better estimator of σ_0^2 than $\bar{\bar{X}}$ is of μ_0 since S_p^2 is computed from the within-sample variation exclusive of the between sample variation.

Another theoretical control chart for dispersion had control limits given by Equations (51) through (53). Let $S_p^* = (1/k) \sum_{h=1}^{k} S_h$, where $E(S_h) = \sigma_h c_2'$ for $h = 1,2,\dots,k$. If $\sigma_1 = \sigma_2 = \dots = \sigma_k = \sigma_0$, then $E(S_p^*) = \sigma_0 c_2'$ and an unbiased estimate of σ_0 would be s_p^*/c_2'. Thus, it seems reasonable that an empirical control chart for S_h should have limits given by

$$UCL = s_p^*\big(1+(3/c_2')(1-c_2'^2)^{1/2}\big) \tag{54}$$

$$CL = s_p^* \tag{55}$$

$$LCL = s_p^*\left(1-(3/c_2')(1-c_2'^2)^{1/2}\right). \tag{56}$$

Let $V_p^* = (1/k)\sum_{h=1}^{k} V_h$, where $V_h = \sqrt{V_h^2}$ and $V_h^2 = (1/n)\sum_{j=1}^{n}(X_{hj}-\bar{X}_h)^2$. Then $E(V_p^*) = \sigma_0 c_2$, and an unbiased estimate of σ_0 would be v_p^*/c_2. Thus, reasonable upper and lower control limits for V_h are given by

$$v_p^*\left(1\pm 3(c_2^2n)^{-1/2}(n-1-nc_2^2)^{1/2}\right).$$

The factors $B_3 = 1 - 3(c_2^2n)^{-1/2}(n-1-nc_2^2)^{1/2}$ and $B_4 = 1 + 3(c_2^2n)^{-1/2}(n-1-nc_2^2)^{1/2}$ are given in Duncan (18) for $n = 2(1)25$. In both of these charts, the number of subgroups is disregarded, although a rule of thumb is that the number of subgroups should be at least 25 with at least 4 or 5 observations per subgroup.

Determining the control of the process variability could also be viewed as a hypothesis testing problem. One would set up the null hypothesis H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$ against all possible alternatives. Specifically, Ω and ω are those subsets of Euclidean 2k space such that

$$\Omega = \{(\mu_1,\ldots,\mu_k,\sigma_1^2,\ldots,\sigma_k^2)\colon -\infty<\mu_h<\infty \text{ and } \sigma_h^2>0,\ h = 1,2,\ldots,k\}$$

and

$$\omega = \{(\mu_1,\ldots,\mu_k,\sigma_1^2,\ldots,\sigma_k^2)\colon -\infty<\mu_h<\infty,\ \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2>0\}.$$

Let σ^2 denote the common but unknown value of $\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2$. Then

$$L(\omega) = \prod_{i=1}^{k} (2\pi\sigma^2)^{-n/2}\exp\left(-(1/2\sigma^2)\sum_{j=1}^{n}(x_{ij}-\mu_i)^2\right)$$

and

$$L(\Omega) = \prod_{i=1}^{k} (2\pi\sigma_i^2)^{-n/2}\exp\left(-(1/2\sigma_i^2)\sum_{j=1}^{n}(x_{ij}-\mu_i)^2\right).$$

Since the logarithms of $L(\omega)$ and $L(\Omega)$ are the sums of k terms which can be maximized separately, it immediately follows that, for $h = 1,2,\ldots,k$, $\hat{\mu}_{h\omega} = \bar{x}_h$, $\hat{\sigma}_\omega^2 = v_p^2$, $\hat{\mu}_{h\Omega} = \bar{x}_h$, $\hat{\sigma}_{h\Omega}^2 = v_h^2$,

$$L(\hat{\omega}) = (2\pi e)^{-nk/2}(v_p^2)^{-nk/2},$$

and

$$L(\hat{\Omega}) = (2\pi e)^{-nk/2}\prod_{h=1}^{k}(v_h^2)^{-n/2}.$$

Note that $V_h^2 = (1/n)\sum_{j=1}^{n}(X_{hj}-\bar{X}_h)^2$ and $V_p^2 = (1/k)\sum_{h=1}^{k}V_h^2$. Thus,

$$\lambda(\underset{\sim}{x}') = \prod_{h=1}^{k}(v_h^2/v_p^2)^{n/2}, \tag{57}$$

where $\underset{\sim}{x}' = [x_{11},\ldots,x_{1n},x_{21},\ldots,x_{2n},\ldots,x_{k1},\ldots,x_{kn}]$. Bartlett (4) considered a modification of $\lambda(\underset{\sim}{x}')$, denoted $\lambda^*(\underset{\sim}{x}')$, in which the sample sizes are replaced by the degrees of freedom. Thus,

$$\lambda^*(\underset{\sim}{x}') = \prod_{h=1}^{k} (s_h^2/s_p^2)^{(n-1)/2}. \tag{58}$$

If H_0 is true, the asymptotic distribution of either $-2\ln\lambda(\underset{\sim}{x}')$ or $-2\ln\lambda^*(\underset{\sim}{x}')$ is $\chi^2(k-1)$. This asymptotic approximation can be improved by considering the distribution of $-2c\ln\lambda^*(\underset{\sim}{x}')$, where the constant c is chosen so that $E\left[-2c\ln\lambda^*(\underset{\sim}{x}')\right] = (k-1) + O\left((n-1)^{-3}\right)$, where $\lim_{n\to\infty} O\left((n-1)^{-3}\right)/(n-1)^{-3} = M$, a bounded constant. Kendall and Stuart (47) present a detailed derivation of c for $-2c\ln\lambda(\underset{\sim}{x}')$. For $-2c\ln\lambda^*(\underset{\sim}{x}')$, Bartlett has shown that $c^{-1} = 1 + \left[(k+1)/(3kn-3k)\right]$. Using Bartlett's procedure, the form of the critical region is given by

$$w = \{\underset{\sim}{x}' \colon c\left[k(n-1)\ln s_p^2 - (n-1)\sum_{h=1}^{k} \ln s_h^2\right] > \chi^2(k-1,\alpha)\}.$$

When the sample sizes of the subpopulations are identical, the tests based on $\lambda(\underset{\sim}{x}')$ and $\lambda^*(\underset{\sim}{x}')$ are equivalent since $kn\ln\lambda^*(\underset{\sim}{x}') = k(n-1)\ln\lambda(\underset{\sim}{x}')$. Box (10) has demonstrated that this test is very sensitive to nonnormality of the k populations, and he has proposed an approximate test which can be treated as a one-way analysis of variance.

If the null hypothesis is rejected using Bartlett's test, the decision maker would be interested in obtaining confidence intervals for each of the k subgroup population variances. For h = 1,2,...,k, let $\underline{\sigma}_h^2 = (n-1)S_h^2/\chi^2(n-1,\alpha/2k)$, $\bar{\sigma}_h^2 = (n-1)S_h^2/\chi^2(n-1,1-\alpha/2k)$, and $(\underline{\sigma}_h^2, \bar{\sigma}_h^2)$ be $100\left[1-(\alpha/k)\right]\%$ confidence intervals for σ_h^2. If A_h denotes the event that $(\underline{\sigma}_h^2, \bar{\sigma}_h^2)$ covers σ_h^2, then $P\left(\bigcap_{h=1}^{k} A_h\right) \geq 1 - \alpha$ by Bonferroni's inequality.

Empirical dispersion charts can also be based upon the sample range. The theoretical control chart for R had upper and lower limits given by $\sigma_0(d_2 \pm 3d_3)$ with a central line given by $\sigma_0 d_2$. Let $\bar{R} = (1/k)\sum_{h=1}^{k} R_h$. Since $E(\bar{R}) = \sigma_0 d_2$, reasonable control limits for R_h, $h = 1,2,\ldots,k$, would be $\bar{r}\left[1 \pm 3(d_3/d_2)\right]$ with a central line given by $\bar{r}$. Tables of $D_3 = 1 - 3(d_3/d_2)$ and $D_4 = 1 + 3(d_3/d_2)$ are given in Bowker and Lieberman (8) for $n = 2(1)25$. Using percentage points of W, reasonable upper and lower control limits for R_h are $(\bar{r}/d_2)w_{\alpha/2,n}$ and $(\bar{r}/d_2)w_{1-\alpha/2,n}$, respectively.

The greatest limitation of the empirical dispersion charts is that unbiasedness is the only criterion used in replacing a parameter by its estimate, while the number of subgroups is completely ignored. The hypothesis testing viewpoint overcomes this, but, unfortunately, this viewpoint precludes a control chart environment. Another way to account for the number of subgroups is by using the distribution of the average range, $\bar{R}$, whose approximations have been thoroughly investigated and are synopsized by H. David (16). Unfortunately, the approximate distributions are dependent upon σ_0. To simultaneously preserve the control chart setting and take into account the number of subgroups, the following alternative is suggested.

Assume $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma_0^2$. Then $k(n-1)S_p^2/\sigma_0^2$ is distributed $\chi^2\left[k(n-1)\right]$ and $(n-1)S_h^2/\sigma_0^2$ is distributed $\chi^2(n-1)$ for $h = 1,2,\ldots,k$. Now $(n-1)kS_p^2 = (n-1)(S_1^2+\ldots+S_{h-1}^2+S_{h+1}^2+\ldots+S_k^2) + (n-1)S_h^2 = X_{k-h} + X_h$, and $k(n-1)S_p^2/\sigma_0^2 = (X_{k-h}/\sigma_0^2) + (X_h/\sigma_0^2) = \chi^2\left[(k-1)(n-1)\right] + \chi^2(n-1)$. If Y denotes a gamma random variable with parameters α and β, then $f_Y(y) = \left[\Gamma(\alpha)\beta^{\alpha}\right]^{-1} y^{\alpha-1} e^{-y/\beta}$, $y > 0$, and a $\chi^2(\nu)$ is a special case of a Gamma random variable with $\alpha = \nu/2$ and $\beta = 2$. From the independence of X_{k-h} and X_h and the reproductive property of the Gamma, it immediately follows that (X_{k-h}/σ_0^2) is Gamma with $\alpha = (k-1)(n-1)/2$, $\beta = 2$, (X_h/σ_0^2) is Gamma with $\alpha = (n-1)/2$, $\beta = 2$, and $(X_{k-h}/\sigma_0^2) + (X_h/\sigma_0^2)$ is Gamma with $\alpha = k(n-1)/2$ and $\beta = 2$. An additional property of the Gamma is that if Y_1 and Y_2 are independent Gamma random variables with parameters α_1, β and α_2, β, respectively, then $U = Y_1/(Y_1+Y_2)$ has a Beta distribution with parameters α_1 and α_2. That is, $f_U(u) = \{B(\alpha_1,\alpha_2)\}^{-1}u^{\alpha_1-1}(1-u)^{\alpha_2-1}$, $0 \le u \le 1$. Specifically, $Y_1 = (X_h/\sigma_0^2)$ and $Y_2 = (X_{k-h}/\sigma_0^2)$ with $\alpha_1 = (n-1)/2$ and $\alpha_2 = (k-1)(n-1)/2$. If $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma_0^2$, then $S_h^2/kS_p^2 = U$ has a Beta distribution with parameters α_1 and α_2 for $h = 1,2,\ldots,k$. Thus, if $b_{1-\alpha}$ and b_α denote the lower and upper alpha percentage points, respectively, of the Beta distribution, then $P(b_{1-\alpha/2} \le S_h^2/kS_p^2 \le b_{\alpha/2}) = 1 - \alpha$. Equivalently, $P(kb_{1-\alpha/2} \le S_h^2/S_p^2 \le kb_{\alpha/2}) = 1 - \alpha$. In view of this, reasonable upper and lower control limits for S_h^2 would be $s_p^2 kb_{\alpha/2}$ and $s_p^2 kb_{1-\alpha/2}$, respectively. These limits could have also been presented using percentage points of the F distribution since $2\alpha_2 U/(2\alpha_1 - 2\alpha_1 U)$ has an F distribution with $\nu_1 = 2\alpha_1$ and $\nu_2 = 2\alpha_2$. The biggest advantage of the control limits $s_p^2 kb_{\alpha/2}$ and $s_p^2 kb_{1-\alpha/2}$ is that they are directly dependent upon the number of subgroups used to estimate σ_0^2. To investigate the properties of the statistics S_h^2/kS_p^2, it would be useful to have the joint distribution of these ratios. Since these ratios are dependent, this is an avenue of further research. However, a bound can be obtained which is sharper than the Bonferroni bound. Let $U_h = S_h^2/kS_p^2$, for $h = 1,2,\ldots,k$. Let $F_{U_1,\ldots,U_k}(u_1,\ldots,u_k)$ denote $P(U_1 \le u_1,\ldots,U_k \le u_k)$.

Esary et al. (21) have shown that for any k non-negative increasing functions $g_h(X)$ of a random variable X, $E\left(\prod_{h=1}^{k} g_h(x)\right) \geq \prod_{h=1}^{k} E\left(g_h(X)\right)$. This inequality is reiterated and applied by H. David (16) to a similar situation. Let $g_h(x) = P(S_h^2 \leq x)$. Then, by the properties of the conditional distribution function,

$$F_{U_1,\ldots,U_k}(u_1,\ldots,u_k) = \int_0^\infty F_{[S_1^2,\ldots,S_k^2]|S_p^2}(s_p^2 ku_1,\ldots,s_p^2 ku_k) f_{S_p^2}(s_p^2) ds_p^2$$

$$= \int_0^\infty \left(\prod_{h=1}^{k} F_{S_h^2|S_p^2}(s_p^2 ku_h)\right) f_{S_p^2}(s_p^2) ds_p^2$$

$$= E\left(\prod_{h=1}^{k} g_h(S_p^2 ku_h)\right) \geq \prod_{h=1}^{k} E\left(g_h^2(S_p^2 ku_h)\right)$$

$$= \prod_{h=1}^{k} F_{U_h}(u_h).$$

Since $P(U_h > b_{\alpha/k}) = \alpha/k$ and if the total α is allocated equally, then

$$F_{U_1,\ldots,U_k}(b_{\alpha/k},\ldots,b_{\alpha/k}) \geq \left(1-(\alpha/k)\right)^k.$$

This bound is sharper than the Bonferroni bound since $\left(1-(\alpha/k)\right)^k \geq 1 - \alpha$, where $1 - \alpha$ is the Bonferroni bound.

This concludes the univariate dispersion charts.

2. Multivariate Dispersion Control Charts

Theoretical Charts

It is desired that the process remain at the nominal value Σ_0. To check this, a random sample of size n is obtained and the value of some sample statistic is determined from the $(n\times p)$ data matrix X.

The first chart to be considered is the multivariate analogue of the S^2 chart, which was also viewed as a hypothesis testing problem. The decision maker would be interested in testing H_0: $\Sigma = \Sigma_0$ vs. H_1: $\Sigma \neq \Sigma_0$ based on a random sample of size n from the population. Specifically, Ω and ω are those subsets of Euclidean $(p^2+3p)/2$ space such that

$$\Omega = \{(\underset{\sim}{\mu},\Sigma)\colon -\infty<\underset{\sim}{\mu}<\infty,\ \Sigma \text{ is positive definite}\}$$

and

$$\omega = \{(\underset{\sim}{\mu},\Sigma)\colon -\infty<\underset{\sim}{\mu}<\infty,\ \Sigma = \Sigma_0\}.$$

It immediately follows from Theorem 3.1 that $\hat{\underset{\sim}{\mu}}_\Omega = \bar{\underset{\sim}{x}}$, $\hat{\Sigma}_\Omega = n^{-1}A$, and

$$L(\hat{\Omega}) = (2\pi/n)^{-(pn)/2}|A|^{-n/2}e^{-(np)/2}.$$

By using a procedure similar to that used in Theorem 3.1, it follows that $\hat{\underset{\sim}{\mu}}_\omega = \bar{\underset{\sim}{x}}$, $\hat{\Sigma}_\omega = \Sigma_0$, and

$$L(\hat{\omega}) = (2\pi)^{-(pn)/2}|\Sigma_0|^{-n/2}e^{-(1/2)\mathrm{tr}(\Sigma_0^{-1}A)}.$$

Thus, if X denotes the $(p\times n)$ data matrix, then

$$\lambda(X) = (e/n)^{(pn)/2}|\Sigma_0^{-1}A|^{n/2}\exp\{-(1/2)\mathrm{tr}(\Sigma_0^{-1}A)\}. \tag{59}$$

This result is stated in Anderson (3). For p = 1, this reduces to the result derived earlier. Let λ denote $\lambda(X)$. Then, using a procedure similar to Anderson, it can be shown that under the null hypothesis,

$$E(\lambda^r) = (2e/n)^{pnr/2}(1+r)^{-(p/2)(nr+n-1)} \prod_{h=1}^{p} \Gamma\{(nr+n-h)/2\}/\Gamma\{(n-h)/2\}.$$

Since $\lambda^{-2it} = \exp(-2it\ell n\lambda)$, it follows that the characteristic function of $-2\ell n\lambda$ is $E(\lambda^{-2it})$, where

$$E(\lambda^{-2it}) = (2e/n)^{-ipnt}(1-2it)^{-(p/2)(n-1-2int)} \prod_{h=1}^{p} \Gamma\{(n-h-2int)/2\}/\Gamma\{(n-h)/2\}.$$

This result is valid for all real t only when the above gamma functions exist. Anderson then shows that $-2\ell n\lambda$ is asymptotically distributed as $\chi^2\{p(p+1)/2\}$. This is the usual asymptotic likelihood ratio test result. Thus, to test H_0: $\Sigma = \Sigma_0$ vs. H_1: $\Sigma \neq \Sigma_0$, the decision maker would select a random sample of size n, compute

$$-2\ell n\lambda = -pn + pn\ell nn - n\ell n(|A|/|\Sigma_0|) + \mathrm{tr}(\Sigma_0^{-1}A), \tag{60}$$

and reject H_0 whenever $-2\ell n\lambda$ exceeds $\chi^2\{p(p+1)/2,\alpha\}$. Thus, the asymptotic upper control limit is given by $\chi^2\{p(p+1)/2,\alpha\}$. Korin (48) showed that the distribution of $-2\ell n\lambda$, slightly modified, may be represented as a series of central χ^2 distributions. By limiting the series to 15 terms, he found by empirical observation for selected n, p, and α = .01 and .05 that the likelihood asymptotic χ^2 and series percentage points are quite close even for moderate n. He also proposes an F approximation which appears to be better. Finally, he gives the .01 and .05 percentage points of $-2\ell n\lambda$, slightly modified, using the 15-term series approximation for p = 2(1)10 and selected values of n. The power of testing H_0: $\Sigma = \Sigma_0$ vs. H_1: $\Sigma \neq \Sigma_0$ can also be approximated by using asymptotic results of the likelihood ratio test criterion.

Examination of Equation (60) reveals that the test for dispersion is partly based on $|A|$. This suggests that $|A|$ is a univariate measure of multivariate dispersion. Since $S = (n-1)^{-1}A$, the same holds true for the determinant of S.

To gain some insight into $|A|$ as a measure of dispersion, consider p = 2 and a sample of size n. The data will not be considered in raw form but as deviations from the mean. That is,

$$D = X - \underset{\sim}{\bar{x}}\underset{\sim}{j}_n' = \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_1 & \cdots & x_{1n}-\bar{x}_1 \\ x_{21}-\bar{x}_2 & x_{22}-\bar{x}_2 & \cdots & x_{2n}-\bar{x}_2 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{d}_1' \\ \underset{\sim}{d}_2' \end{bmatrix},$$

where $\underset{\sim}{j}_n' = [1,1,\ldots,1]$. Note that $A = DD'$ and $S = (n-1)^{-1}DD'$. Recall that the maximum likelihood estimate of the correlation coefficient, denoted by r, is $r = \underset{\sim}{d}_1'\underset{\sim}{d}_2/\|d_1\|\|d_2\|$ and that r equals the cosine of the angle between $\underset{\sim}{d}_1$ and $\underset{\sim}{d}_2$. Thus

$$A = \begin{bmatrix} \|\underset{\sim}{d}_1\|^2 & \underset{\sim}{d}_1'\underset{\sim}{d}_2 \\ \underset{\sim}{d}_1'\underset{\sim}{d}_2 & \|\underset{\sim}{d}_2\|^2 \end{bmatrix} = \begin{bmatrix} (n-1)s_1^2 & (n-1)rs_1s_2 \\ (n-1)rs_1s_2 & (n-1)s_2^2 \end{bmatrix},$$

and $|A| = (n-1)^2s_1^2s_2^2(1-r^2) = (n-1)^2s_1^2s_2^2\sin^2\theta$. But $(n-1)s_1s_2\sin\theta = \|\underset{\sim}{d}_1\|\,\|\underset{\sim}{d}_2\|\sin\theta$, which is the area of the parallelogram formed by using $\underset{\sim}{d}_1$ and $\underset{\sim}{d}_2$ as principal edges. This is illustrated in Figure 7.

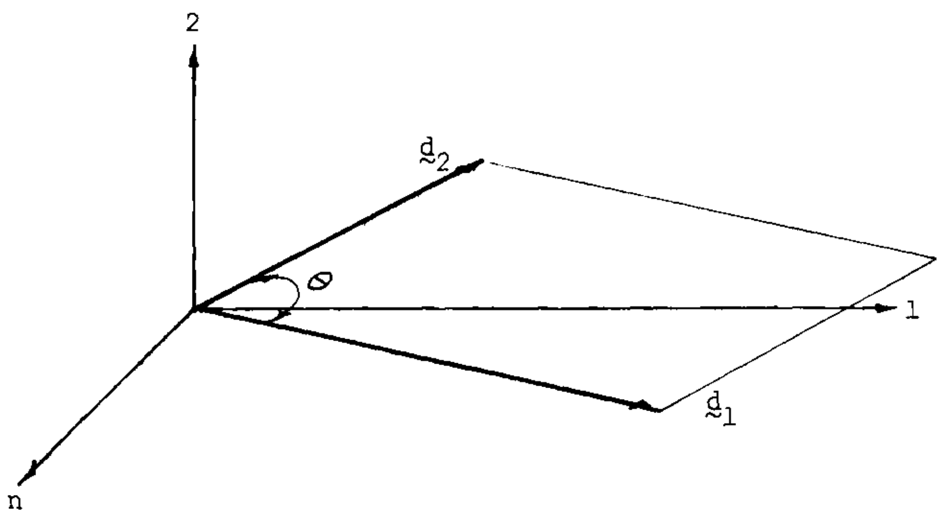

Figure 7. The Square Root of the Area of the Sample Generalized Variance

Thus, $|A| = |DD'|$ is the square of this area. Anderson (3) generalizes this result for general p. That is, $|DD'| = |A| = (n-1)^p|S|$ is the square of the p-dimensional volume of the parallelotope which has $\underset{\sim}{d}_1, \underset{\sim}{d}_2, \ldots, \underset{\sim}{d}_p$ as principal edges. The quantity, $|S|$, is called the sample generalized variance. One disadvantage in using $|S|$ as a measure of dispersion is that it is subject to all the properties of determinants, which implies that different matrices give the same value of $|S|$. Thus, the process dispersion may go astray but remain undetected.

The distributional properties of $|A|$ are found with the aid of Equation (8). Assume $\Sigma = \Sigma_0$, and let $\{c(p,n-1)\}^{-1} = 2^{(n-1)p/2}\pi^{p(p-1)/4}$ $|\Sigma_0|^{(n-1)/2} \prod_{h=1}^{p} \Gamma\{(n-h)/2\}$. Then $E(|A|^r) = c(p,n-1) \int \cdots$ $\int |A|^{(n-1+2r-p-1)/2} e^{-(1/2)\mathrm{tr}\Sigma_0^{-1}A}\, dA$, where the region of integration is over the permissible values of A and dA denotes $\prod_{i\le j}^{p} da_{ij}$. Since the above multiple integral equals $\{c(p,n-1+2r)\}^{-1}$, it immediately follows that $E(|A|^r) = c(p,n-1)/c(p,n-1+2r)$. This simplifies to yield

$$E(|A|^r) = 2^{rp}|\Sigma_0|^r \prod_{h=1}^{p} \Gamma\{(n-h)/2 + r\}/\Gamma\{(n-h)/2\}. \tag{61}$$

Let Y_ν denote $\chi^2(\nu)$ and let $U = \prod_{h=1}^{p} Y_{n-h}$, where the Y's are all independent. Then $E(U^r) = \prod_{h=1}^{p} E(Y_{n-h}^r) = \prod_{h=1}^{p} 2^r\Gamma\{(n-h)/2 + r\}/\Gamma\{(n-h)/2\}$. A set of moments determines the distribution uniquely if $\sum_{j=0}^{\infty} (\nu_j t^j)/j!$ converges for some real nonzero t, where ν_j denotes the j*th* absolute moment. A corollary of this is that if the limit as $r \to \infty$ of $(\nu_r^{1/r})/r$ is finite then the distribution is uniquely determined. Since this condition is satisfied, it follows that

$$|A| = |\Sigma_0| \prod_{h=1}^{p} \chi^2_{n-h} \tag{62}$$

and

$$|S| = (n-1)^{-p}|\Sigma_0| \prod_{h=1}^{p} \chi^2_{n-h}, \tag{63}$$

where the chi-square random variables are independent. The mean and variance of $|A|$ are determined from Equation (61). Thus,

$$E(|A|) = |\Sigma_0| \prod_{h=1}^{p} (n-h) \tag{64}$$

and

$$V(|A|) = |\Sigma_0|^2 \prod_{i=1}^{p} (n-i)[\prod_{j=1}^{p} (n-j+2) - \prod_{j=1}^{p} (n-j)] \tag{65}$$

Equation (63) is of little use in applications since the distribution of the product of independent chi-square random variables is not known for $p > 2$. However, Anderson (3) proves that $\sqrt{n-1}\ \left((|S|/|\Sigma_0|)-1\right)$ is asymptotically normally distributed with mean 0 and variance 2p based on a random sample of size n from $N(\mu,\Sigma_0)$. Anderson's approximation improves with increasing n.

Hoel (33) has suggested another approximation. Let $U = \prod_{h=1}^{p} Y_{n-h}$, where the Y's are all independent chi-squared random variables with n - h degrees of freedom. Hoel suggests that $U^{1/p}$ is approximately distributed as a Gamma random variable with parameters $\alpha = p(n-p)/2$ and $\beta^{-1} = (p/2)[1-(p-1)(p-2)/2n]^{1/p}$. This distribution is exact for p = 1 and p = 2. Thus, for p = 1, $U = (n-1)S^2/\sigma_0^2$ is distributed as a Gamma random variable with $\alpha = (n-1)/2$ and $\beta = 2$, which is a chi-squared random variable with (n-1) degrees of freedom. This is the familiar univariate result. For p = 2, $U^{1/2} = (n-1)|S|^{1/2}/|\Sigma_0|^{1/2}$ is distributed as a Gamma random variable with $\alpha = (2n-4)/2$ and $\beta = 1$. Thus, for $p = 2$, $2U^{1/2}$ is distributed as a chi-squared random variable with $(2n-4)$ degrees of freedom. Gnanadesikan and Gupta (24) have performed an empirical investigation of Hoel's approximation for selected p and n. For $p = 3$ and $n = 5$ and 15, the approximation appears to be quite good. For $p = 5$ and 10 with $n = 15$ and 25, the accuracy of Hoel's approximation appears to decrease for increasing p.

A final approximation is based on $(1/p)\ell n U = (1/p) \sum_{h=1}^{p} \ell n(Y_{n-h})$. Let γ_1 and γ_2 denote the coefficients of skewness and kurtosis, respectively, where $\gamma_1 = \mu_3/\mu_2^{3/2}$, $\gamma_2 = (\mu_4/\mu_2^2) - 3$, and μ_r denotes the rth central moment. If $X \sim N(\mu,\sigma^2)$, then $\gamma_1 = \gamma_2 = 0$. If X denotes a Gamma random variable with parameters α and β, then $\gamma_1 = 2/\sqrt{\alpha}$ and $\gamma_2 = 6/\alpha$. Since $\chi^2(\nu)$ is a special case of a Gamma random variable with $\alpha = \nu/2$ and $\beta = 2$, it follows that, for $\chi^2(\nu)$, $\gamma_1 = 2^{3/2}/\sqrt{\nu}$ and $\gamma_2 = 12/\nu$, where $\gamma_1 = \gamma_2 = 0$ as $\nu \to \infty$. Now Bartlett and Kendall (6) have suggested using the $\ell n\chi^2(\nu)$ in place of $\chi^2(\nu)$. Let $\kappa(t)$ and κ_r denote the cumulant generating function and the rth cumulant, respectively. Also, let $Y = \ell nX$ where X is Gamma distributed with α and β. Then $M_Y(t) = \Gamma(\alpha+t)\beta^t/\Gamma(\alpha)$ and $\kappa(t) = \ell n M_Y(t) = \ell n\Gamma(\alpha+t) + t\ell n\beta - \ell n\Gamma(\alpha)$. Recall that the digamma function, denoted by ψ, is defined to be such that $\psi(z) = d[\ell n\Gamma(z)]/dz = \Gamma'(z)/\Gamma(z)$. The polygamma function, denoted by $\psi^{(h)}$ for $h = 1,2,\ldots,$ is defined to be such that $\psi^{(h)}(z) = d^h\psi(z)/dz^h$. Tables of $\psi(z)$ and $\psi^{(1)}(z)$ are contained in Abramowitz and Stegun (1), together with properties of $\psi^{(h)}(z)$. More extensive tables are contained in Davis (17). Johnson and Kotz (44) state that a very good approximate formula for $z \geq 2$ is $\psi(z) \approx \ln\left(z-(1/2)\right)$. Thus, if X is a Gamma random variable with parameters α and β and $Y = \ln X$, then $\kappa_1(Y) = \psi(\alpha) + \ln\beta$ and $\kappa_r(Y) = \psi^{(r-1)}(\alpha)$ for $r \geq 2$. Since $\chi^2(\nu)$ is Gamma distributed with $\alpha = \gamma/2$ and $\beta = 2$, it follows that $\kappa_1\left(\ln\chi^2(\nu)\right) = \psi(\nu/2) + \ln 2$ and $\kappa_r\left[\ln\chi^2(\nu)\right] = \psi^{(r-1)}(\nu/2)$ for $r \geq 2$. In general, $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and $\gamma_2 = \kappa_4/\kappa_2^2$. Thus, for $Y_\nu = \ln\chi^2(\nu)$, $\gamma_1 \approx -\sqrt{2}/\sqrt{\nu-1}$ and $\gamma_2 \approx 4(\nu-1)^{-1}$. Since $2\sqrt{2}/\sqrt{\nu} > \sqrt{2}/\sqrt{\nu-1}$ for $\nu > (4/3)$ and $12\nu^{-1} > 4(\nu-1)^{-1}$ for $\nu > (3/2)$, the $\ln\chi^2(\nu)$ is more nearly normally distributed than $\chi^2(\nu)$. Bartlett and Kendall suggest that for $\nu > 9$ the normal distribution provides a fairly good approximation to the distribution of $\ln\chi^2(\nu)$, and the normal approximation is better for $\ln\chi^2(\nu)$ than $\chi^2(\nu)$. Thus, this normal approximation to $\ln\chi^2(\nu)$ in conjunction with the Central-Limit Theorem asserts that $\ln U = \sum_{h=1}^{p} \ln(Y_{n-h})$ is approximately normally distributed with an improvement in the approximation for both increasing n and p. Gnanadesikan and Gupta represent $(1/p)\ln U$ as a Type A (Gram-Charlier) series to empirically investigate the approximation for increasing p, and they have found that there is considerable improvement for increasing p. Now $E(\ln U) = \sum_{h=1}^{p} E\left(\ln(Y_{n-h})\right) = p\ln 2 + \sum_{h=1}^{p} \psi\left((n-h)/2\right)$, since $\kappa_1 = \mu_1'$. Also, $V(\ln U) = \sum_{h=1}^{p} V\left(\ln(Y_{n-h})\right) = \sum_{h=1}^{p} \psi^{(1)}\left[(n-h)/2\right]$, since $\kappa_2 = \mu_2$. Thus

$$\ln(|A|/|\Sigma_0|) \approx N\left\{p\ln 2+\sum_{h=1}^{p}\psi\left((n-h)/2\right), \sum_{h=1}^{p}\psi^{(1)}\left((n-h)/2\right)\right\}. \tag{66}$$

Since $|A| = (n-1)^p|S|$, this approximate normality also holds for

$\ell n\{(n-1)^p|S|/|\Sigma_0|\}$. For most cases, the value of the mean could be obtained from the digamma tables. For non-tabled entries or in the absence of tables, the following formulas may be of benefit for integer and half-integer values:

$$\psi(n) = -\gamma + \sum_{h=1}^{n-1} h^{-1}, \text{ for } n \geq 2,$$

and

$$(1/2)\psi\{n+(1/2)\} = \psi(2n) - (1/2)\psi(n) - \ell n 2,$$

where

$$\gamma = .57721\ 56649\ \ldots\ .$$

The value of the variance can be obtained by using the following properties of the trigamma function:

$$\psi^{(1)}(n+1) = (\pi^2/6) - \sum_{h=1}^{n} h^{-2}$$

and

$$\psi^{(1)}\{n+(1/2)\} = (\pi^2/2) - 4 \sum_{h=1}^{n} (2h-1)^{-2}.$$

Approximations to the mean and variance could be obtained by using Johnson and Kotz's approximation. Thus, $E(|A|/|\Sigma_0|) \approx \sum_{h=1}^{p} \ell n(n-1-h)$ and $V(|A|/|\Sigma_0|) \approx \sum_{h=1}^{p} 2(n-1-h)^{-1}$.

The first theoretical control chart for dispersion was viewed as a hypothesis testing problem. The other theoretical charts to be considered are more heuristic. The first of these is the multivariate analogue of the univariate sigma chart.

Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$ be a random sample from $N(\underset{\sim}{\mu}, \Sigma_0)$. Since most of the probability distribution of $|S|$ is contained in the interval $E(|S|) \pm 3\sqrt{V(|S|)}$, it seems reasonable that a control chart for $|S|$ should have $E(|S|) + 3\sqrt{V(|S|)}$ and $E(|S|) - 3\sqrt{V(|S|)}$ as its upper and lower limits, respectively, with a center line given by $E(|S|)$. From Equations (64) and (65), it immediately follows that

$$E(|S|) = |\Sigma_0|(n-1)^{-p} \prod_{h=1}^{p} (n-h) = |\Sigma_0| b_1$$

and

$$V(|S|) = |\Sigma_0|^2 (n-1)^{-2p} \prod_{i=1}^{p} (n-i)\left[\prod_{j=1}^{p} (n-j+2) - \prod_{j=1}^{p} (n-j)\right] = |\Sigma_0|^2 b_2$$

For $p = 1$, these equations become $E(S^2) = \sigma_0^2$ and $V(S^2) = 2\sigma_0^4(n-1)^{-1}$. For general p, one control chart for $|S|$ would have control limits given by

$$\text{UCL} = |\Sigma_0|(b_1 + 3b_2^{1/2}) \tag{67}$$

$$\text{CL} = |\Sigma_0| b_1 \tag{68}$$

$$\text{LCL} = |\Sigma_0|(b_1 - 3b_2^{1/2}), \tag{69}$$

where

$$b_1 = (n-1)^{-p} \prod_{h=1}^{p} (n-h)$$

and

$$b_2 = (n-1)^{-2p} \prod_{i=1}^{p} (n-i)[\prod_{j=1}^{p} (n-j+2) - \prod_{j=1}^{p} (n-j)].$$

As in the univariate case, these limits cannot be thought of as .9973 probability limits. Note that these charts are based on the sample variances and covariances and not the square roots of these quantities. To conform to the univariate sigma chart, consider a control chart for dispersion based on $|S|^{1/2}$. From Equations (61) and (64), it follows that

$$E(|S|^{1/2}) = |\Sigma_0|^{1/2}(n-1)^{-p/2}2^{p/2} \prod_{h=1}^{p} \Gamma\big((n-h+1)/2\big)/\Gamma\big((n-h)/2\big) = |\Sigma_0|^{1/2}b_3,$$

and

$$V(|S|^{1/2}) = |\Sigma_0|(n-1)^{-p}[\prod_{h=1}^{p} (n-h) - 2^p\{\prod_{h=1}^{p} \Gamma\big((n-h+1)/2\big)/\Gamma\big((n-h)/2\big)\}^2]$$

$$= |\Sigma_0|b_4.$$

Since most of the probability distribution of $|S|^{1/2}$ lies within three standard deviations of its mean, the control limits for $|S|^{1/2}$ are given by

$$UCL = |\Sigma_0|^{1/2}(b_3+3b_4^{1/2}) \tag{70}$$

$$CL = |\Sigma_0|^{1/2}b_3 \tag{71}$$

$$LCL = |\Sigma_0|^{1/2}(b_3-3b_4^{1/2}), \tag{72}$$

where

$$b_3 = (n-1)^{-p/2}2^{p/2}\prod_{h=1}^{p}\Gamma\{(n-h+1)/2\}/\Gamma\{(n-h)/2\}$$

and

$$b_4 = (n-1)^{-p}[\prod_{h=1}^{p}(n-h)-2^p\{\prod_{h=1}^{p}\Gamma\{(n-h+1)/2\}/\Gamma\{(n-h)/2\}\}^2].$$

For $p = 1$, the control limits given by Equations (70) through (72) equal those given by Equations (51) through (53). The lower control limits given by Equations (69) and (72) could be replaced by 0 if they are negative. The greatest disadvantage of the multivariate sigma chart is that it does not utilize the distribution of the sample statistic but uses only the first two moments. Even though the distribution of $|S|$ is not tractable for $p > 2$, its approximate distributions have been presented and should be used.

The use of Anderson's approximation implies that

$$P[|\Sigma_0|\{1-z_{\alpha/2}\sqrt{2p/(n-1)}\}\leq|S|\leq|\Sigma_0|(1+z_{\alpha/2}\sqrt{2p/(n-1)})] = 1 - \alpha.$$

Thus, a theoretical control chart for $|S|$ would have its limits given by

$$\mathrm{UCL} = |\Sigma_0|\left(1+z_{\alpha/2}\sqrt{2p/(n-1)}\right) \tag{73}$$

and

$$\mathrm{LCL} = |\Sigma_0|\left(1-z_{\alpha/2}\sqrt{2p/(n-1)}\right). \tag{74}$$

If only an upper control limit is used, $z_{\alpha/2}$ is replaced by z_α. For $\alpha = .0027$, $z_{.00135} = 3.0$, and this would yield the usual three-sigma results. A control chart for $|S|^{1/2}$ would have its control limits given by the positive square root of the limits for $|S|$.

Let $\ell n\mathrm{U} = \ell n\left((n-1)^p|S|/|\Sigma_0|\right)$. Then, from Equation (66),

$$P[c_1+E(\ell n\mathrm{U})-z_{\alpha/2}\sqrt{V(\ell n\mathrm{U})}\le \ell n|S|\le c_1+E(\ell n\mathrm{U})+z_{\alpha/2}\sqrt{V(\ell n\mathrm{U})}] = 1-\alpha,$$

where $E(\ell n\mathrm{U}) = p\ell n2 + \sum_{h=1}^{p}\psi\left((n-h)/2\right)$, $V(\ell n\mathrm{U}) = \sum_{h=1}^{p}\psi^{(1)}\left((n-h)/2\right)$ and $c_1 = \ell n|\Sigma_0| - p\ell n(n-1)$. Thus, a theoretical control chart for $\ell n|S|$ would have

$$\mathrm{UCL} = c_1 + E(\ell n\mathrm{U}) + z_{\alpha/2}\sqrt{V(\ell n\mathrm{U})} \tag{75}$$

and

$$\mathrm{LCL} = c_1 + E(\ell n\mathrm{U}) - z_{\alpha/2}\sqrt{V(\ell n\mathrm{U})} \tag{76}$$

where

$$E(\ell n U) = p\ell n 2 + \sum_{h=1}^{p} \psi\big((n-h)/2\big), \quad V(\ell n U) = \sum_{h=1}^{p} \psi^{(1)}\big((n-h)/2\big), \text{ and}$$

$$c_1 = \ell n|\Sigma_0| - p\ell n(n-1).$$

Let the right-hand sides of Equations (75) and (76) be denoted by δ_2 and δ_1, respectively. Then $P(\delta_1 \le \ell n|S| \le \delta_2) = P(e^{\delta_1} \le |S| \le e^{\delta_2}) = 1 - \alpha$. Thus, another theoretical control chart for $|S|$ would have

$$UCL = \exp\{c_1 + E(\ell n U) + z_{\alpha/2}\sqrt{V(\ell n U)}\} \tag{77}$$

and

$$LCL = \exp\{c_1 + E(\ell n U) - z_{\alpha/2}\sqrt{V(\ell n U)}\}. \tag{78}$$

An avenue of further research is to investigate whether the control limits for $|S|$ given by Equations (77) and (78) are superior to those given by Equations (73) and (74) since the former require more effort to calculate.

The control limits given by Equations (73) and (74) and Equations (77) and (78) only make use of approximate distributions of $|S|$. Although this is a reasonable procedure for $p \ge 3$ and large n, it was previously stated that $2(n-1)|S|^{1/2}/|\Sigma_0|^{1/2}$ is exactly distributed as a chi-squared random variable with (2n-4) degrees of freedom when p = 2. Since

$$P[\chi^2(2n-4,1-(\alpha/2)) \le 2(n-1)|S|^{1/2}/|\Sigma_0|^{1/2} \le \chi^2(2n-4,\alpha/2)] = 1 - \alpha,$$

it immediately follows that a theoretical control chart for $|S|^{1/2}$ would have

$$\text{UCL} = |\Sigma_0|^{1/2}\chi^2(2n-4,\alpha/2)/(2(n-1)) \tag{79}$$

and

$$\text{LCL} = |\Sigma_0|^{1/2}\chi^2(2n-4,1-(\alpha/2))/(2(n-1)). \tag{80}$$

A theoretical control chart for dispersion based on $|S|$ would have its control limits given by the square of those given by Equations (79) and (80). That is, for $p = 2$, the exact control limits for $|S|$ would have

$$\text{UCL} = |\Sigma_0|(\chi^2(2n-4,\alpha/2))^2/(4(n-1)^2) \tag{81}$$

and

$$\text{LCL} = |\Sigma_0|(\chi^2(2n-4,1-(\alpha/2))^2/(4(n-1)^2). \tag{82}$$

In all of the control limits which require percentage points of either the chi-squared or normal distribution, the α-level must be determined by the decision maker.

Empirical Charts

Assume that k rational subgroups of n observations each have been collected, where statistical control existed within each subgroup.

Refer to Table 9.

Table 9. Data for $p > 1$

Sample Number	Data Matrices	Subgroup Statistics	Population Parameters
1	$X^{(1)}$	$\bar{\underset{\sim}{x}}^{(1)}, S^{(1)}$	$\underset{\sim}{\mu}^{(1)}, \Sigma^{(1)}$
2	$X^{(2)}$	$\bar{\underset{\sim}{x}}^{(2)}, S^{(2)}$	$\underset{\sim}{\mu}^{(2)}, \Sigma^{(2)}$
⋮	⋮	⋮	⋮
k	$X^{(k)}$	$\bar{\underset{\sim}{x}}^{(k)}, S^{(k)}$	$\underset{\sim}{\mu}^{(k)}, \Sigma^{(k)}$

Note that Table 9 differs from Table 5 in that now it must be determined whether $\Sigma^{(1)} = \Sigma^{(2)} = \ldots = \Sigma^{(k)} = \Sigma_0$, where Σ_0 is the common but unspecified value of the Σ's. Most of the procedures to be presented are ad hoc procedures and are based on the substitution of the sample estimates for the population parameters in the theoretical charts.

For $h = 1,2,\ldots,k$, let $S^{(h)} = (n-1)^{-1}A^{(h)}$. It was stated in Chapter III that $A^{(h)}$ has a Wishart distribution, denoted $w(A^{(h)}|\Sigma^{(h)}, p,n-1)$. It was also demonstrated that $E(S^{(h)}) = \Sigma^{(h)}$. If $\Sigma^{(1)} = \Sigma^{(2)} = \ldots = \Sigma^{(k)} = \Sigma_0$, then $\sum_{h=1}^{k} A^{(h)}$ is $w\left(\sum_{h=1}^{k} A^{(h)}|\Sigma_0,p,k(n-1)\right)$ from the reproductive property of the Wishart and since the $A^{(h)}$ are independent because of rational subgroups. Furthermore, $S_p = (1/k) \sum_{h=1}^{k} S^{(h)}$ is an unbiased estimator of Σ_0. Since most of the multivariate theoretical dispersion charts contained $|\Sigma_0|$ and $|\Sigma_0|^{1/2}$, unbiased estimators of these quantities are also needed. Since

$S_p = \left(k(n-1)\right)^{-1} \sum_{h=1}^{k} A^{(h)}$, it follows that $\left(k(n-1)\right)S_p$ is $w\left(\sum_{h=1}^{k} A^{(h)} | \Sigma_0, p, k(n-1)\right)$. Thus, by using a procedure similar to the one used in the development of Equation (61), it follows that for $A_p = \sum_{h=1}^{k} A^{(h)}$

$$E(|A_p|^r) = 2^{rp}|\Sigma_0|^r \prod_{h=1}^{p} \Gamma\left((kn-k+1-h)/2 + r\right)/\Gamma\left((kn-k+1-h)/2\right), \quad (83)$$

Since $|A_p| = k^p(n-1)^p|S_p|$, it follows by substitution in Equation (83) that

$$E(|S_p|) = |\Sigma_0|k^{-p}(n-1)^{-p} \prod_{h=1}^{p} (kn-k+1-h) = |\Sigma_0|b_5 \quad (84)$$

and

$$E(|S_p|^{1/2}) = |\Sigma_0|^{1/2}k^{-p/2}(n-1)^{-p/2}2^{p/2} \prod_{h=1}^{p} \Gamma\left((kn-k+2-h)/2\right)/\Gamma\left((kn-k+1-h)/2\right)$$

$$= |\Sigma_0|^{1/2}b_6 \quad (85)$$

Thus, an unbiased estimator for $|\Sigma_0|$ is $|S_p|/b_5$ and an unbiased estimator for $|\Sigma_0|^{1/2}$ is $|S_p|^{1/2}/b_6$. In order to obtain multivariate empirical control charts, the univariate procedure will be adopted. That is, replace the population parameters or functions of them by their unbiased estimates.

The first empirical control chart is formed by using Equation (60), which was the test statistic for testing H_0: $\Sigma = \Sigma_0$ vs. H_1: $\Sigma \neq \Sigma_0$. The control limit would remain at $\chi^2\left(p(p+1)/2, \alpha\right)$ while, for $h = 1,2,\ldots,k$, the test statistic becomes

$$-pn + pn\ell n n - n\ell n\left(|A^{(h)}|/(|S_p|b_5^{-1})\right) + tr\left(S_p^{-1}A^{(h)}\right). \qquad (86)$$

The use of Equation (86) is not strongly advocated since it originated as a result of a test on a covariance matrix for a single sample. In fact, the likelihood ratio test for testing H_0: $\Sigma^{(1)} = \Sigma^{(2)} = \ldots = \Sigma^{(k)}$ versus the alternative that they are not all equal has been developed and will be mentioned later.

The second empirical control chart is formed by using Equations (67) through (69). In this case, the test statistic is $|S^{(h)}|$ and the control limits are given by

$$UCL = (|S_p|b_5^{-1})(b_1+3b_2^{1/2}) \qquad (87)$$

$$CL = (|S_p|b_5^{-1})b_1 \qquad (88)$$

$$LCL = (|S_p|b_5^{-1})(b_1-3b_2^{1/2}), \qquad (89)$$

with

$$b_5 = k^{-p}(n-1)^{-p} \prod_{h=1}^{p} (kn-k+1-h).$$

A variation of the control limits given by Equations (67) through (69) was given by Equations (70) through (72). Since these were the limits for the theoretical sigma chart, the corresponding limits for the empirical sigma chart are given by

$$UCL = (|S_p|^{1/2} b_6^{-1})(b_3 + 3b_4^{1/2}) \tag{90}$$

$$CL = (|S_p|^{1/2} b_6^{-1}) b_3 \tag{91}$$

$$LCL = (|S_p|^{1/2} b_6^{-1})(b_3 - 3b_4^{1/2}) \tag{92}$$

where the test statistic is $|S^{(h)}|^{1/2}$ for $h = 1,2,\ldots,k$, and

$$b_6 = k^{-p/2}(n-1)^{-p/2} 2^{p/2} \prod_{h=1}^{p} \Gamma\big((kn-k+2-h)/2\big)/\Gamma\big((kn-k+1-h)/2\big).$$

The other empirical control charts are based on the asymptotic approximations to $|S|$ and functions of it. From Equations (73) and (74), one empirical control chart for $|S^{(h)}|$ would have its limits given by

$$UCL = (|S_p| b_5^{-1})\big(1 + z_{\alpha/2}\sqrt{2p/(n-1)}\big) \tag{93}$$

and

$$LCL = (|S_p| b_5^{-1})\big(1 - z_{\alpha/2}\sqrt{2p/(n-1)}\big) \tag{94}$$

Let $\ell n U = \ell n\big((n-1)^p |S| / |\Sigma_0|\big)$, where its approximate distribution is given by Equation (66). Based on this, theoretical control charts for $|S|$ were given by Equations (77) and (78). Thus, another empirical control chart for $|S^{(h)}|$ would have its limits given by

$$UCL = \exp\{c_2+E(\ell n U)+z_{\alpha/2}\sqrt{V(\ell n U)}\} \tag{95}$$

and

$$LCL = \exp\{c_2+E(\ell n U)-z_{\alpha/2}\sqrt{V(\ell n U)}\}, \tag{96}$$

where $c_2 = \ell n(|S_p|b_5^{-1})-p\ell n(n-1)$. In this instance, an unbiased estimate was not used for $\ell n\ |\Sigma_0|$. Instead, an unbiased estimate of $|\Sigma_0|$ was used. In general, $E(\ell n X) \neq \ell n E(X)$.

A final empirical control chart is based on Hoel's results for $p = 2$. From Equations (81) and (82), an empirical control chart for $|S^{(h)}|$ would have its limits given by

$$UCL = (|S_p|b_5^{-1})\left(\chi^2(2n-4,\alpha/2)\right)^2/\left(4(n-1)^2\right) \tag{97}$$

and

$$LCL = (|S_p|b_5^{-1})\left(\chi^2(2n-4,1-(\alpha/2))\right)^2/\left(4(n-1)^2\right). \tag{98}$$

Note that these limits are valid only for $p = 2$.

As in the univariate case, one of the biggest disadvantages of the empirical dispersion charts is that since unbiasedness is the major criterion it may not be as powerful as other methods.

Determining the control of the process variability could also be viewed as a hypothesis testing problem. One would set up the null hypothesis H_0: $\Sigma^{(1)} = \Sigma^{(2)} = \ldots = \Sigma^{(k)}$ against all possible alternatives. Specifically, Ω and ω are those subsets of Euclidean

$pk + k(p^2+p)/2$ space such that

$$\Omega = \{(\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(k)},\Sigma^{(1)},\ldots,\Sigma^{(k)}): -\infty<\underset{\sim}{\mu}^{(h)}<\infty,\ \Sigma^{(h)} \text{ is positive definite}\}$$

and

$$\omega = \{(\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(k)},\Sigma^{(1)},\ldots,\Sigma^{(k)}): -\infty<\underset{\sim}{\mu}^{(h)}<\infty, \Sigma^{(1)} = \Sigma^{(2)} = \ldots = \Sigma^{(k)}\}.$$

Since the recent text by Kshirsagar (49) gives a very detailed development of the likelihood ratio criterion for this hypothesis, only the final results will be presented. Let λ denote $\lambda(X^{(1)},X^{(2)},\ldots,X^{(k)})$. Then

$$\lambda = \frac{\left[\prod_{h=1}^{k} |A^{(h)}|\right]^{n/2} k^{knp/2}}{|A^{(1)}+A^{(2)}+\ldots+A^{(k)}|^{kn/2}}, \tag{99}$$

and the usual asymptotic likelihood ratio test says to reject H_0 if $-2\ell n\lambda$ exceeds $\chi^2\left[(k-1)(p^2+p)/2,\alpha\right]$. Bartlett (4) considered a modification of λ, denoted λ^*, which except for a numerical constant, is given by

$$\lambda^* = \frac{\left[\prod_{h=1}^{k} |A^{(h)}|\right]^{(n-1)/2}}{|A^{(1)}+A^{(2)}+\ldots+A^{(k)}|^{k(n-1)/2}} \tag{100}$$

Due to a result of Box (9), $-2\rho\ell n(k^{k(n-1)p/2}\lambda^*)$ is approximately distributed $\chi^2\left[(k-1)(p^2+p)/2\right]$ where

$$\rho = 1 - \frac{(k^2-1)(2p^2+3p-1)}{6k(n-1)(p+1)(k-1)} .$$

Thus,

$$w = \{(X^{(1)},X^{(2)},\ldots,X^{(k)}):\ -2\rho\ln(k^{k(n-1)p/2}\lambda^{*})>\chi^2\left[(k-1)(p^2+p)/2\right]\}.$$

Higher order approximations can be obtained if necessary.

Siotani (65,66) has developed extensions of the univariate range. Consider a random sample of size n from a p-variate normal, and let

$$R^2_{MAX} = \max_{i<j}\{\underset{\sim}{X}_i-\underset{\sim}{X}_j)'\Sigma^{-1}(\underset{\sim}{X}_i-\underset{\sim}{X}_j)\}$$

and

$$\mathcal{R}^2_{MAX} = \max_{i<j}\{(\underset{\sim}{X}_i-\underset{\sim}{X}_j)'S^{-1}(\underset{\sim}{X}_i-\underset{\sim}{X}_j)\}.$$

Then Siotani gives a method of obtaining the approximate upper α percentage points for α = .01 and .05. Note that R^2_{MAX} and $\mathcal{R}^2_{MAX}$ reduce to the square of the previously stated univariate results when p = 1.

This chapter commenced with a brief review of previously developed theoretical and empirical univariate control charts and the likelihood ratio tests of significance under both circumstances. When there are k rational subgroups, the statistics S^2_h/kS^2_p, h = 1,2,...,k were proposed as a new alternative, and control limits were found for

S_h^2. Also, a bound was obtained for the joint distribution of the S_h^2/kS_p^2.

For the multivariate case, theoretical charts were considered first. The first of these merely adopted the likelihood ratio methodology for testing H_0: $\Sigma = \Sigma_0$ vs. H_1: $\Sigma \neq \Sigma_0$ based on a random sample of size n. It was then suggested that ad hoc theoretical control charts could be based upon the sample generalized variance and its positive square root. After an extensive literature survey, it seems that the concept of using functions of $|S|$ as measures of dispersion in relation to control charts has not previously appeared in the literature. One theoretical chart for $|S|$ borrowed upon the univariate procedure by looking at $E(|S|) \pm 3\sqrt{V(|S|)}$. The other theoretical charts for $|S|$ were based on the asymptotic approximations of Anderson (3) and Gnanadesikan and Gupta (24) and the exact distribution for p = 2. Finally, empirical control charts were developed based on the various theoretical charts and the procedure of replacing population parameters by their unbiased estimates. It seems as though most of these concepts have not previously appeared in the literature.

CHAPTER VIII

SUMMARY AND RECOMMENDATIONS

One of the major purposes of this study was to elaborate upon and develop multivariate analogues of the univariate control charts for location and dispersion. These various charts are summarized below in Table 10.

The first chart considered control of the mean with Σ known, and this was the basis for Chapter IV. As previously stated, Hotelling (36) first suggested using a χ^2 control chart. Furthermore, it was pointed out that this control chart can be viewed as repeated tests of significance of the form H_0: $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs. H_1: $\underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ with Σ known. Thus, Chapter IV merely elaborated upon the previously developed tests of significance methodology, including the power concept and the use of simultaneous confidence intervals, and placed these concepts in a control chart environment. It was demonstrated how the multivariate chart reduces to the well-known univariate chart. Also, Conjecture 4.1 and Table 3 suggest that the Bonferroni intervals are superior to the Scheffe intervals for $\alpha < .55$ and all $p \geq 2$. The proof of this Conjecture is an avenue of future research. Another topic of future research is the effect of correlations on component intervals.

Chart number three in Table 10 is used for analyzing past data for lack of control of the mean when there are k rational subgroups. This was the basis for the first part of Chapter V.

Table 10. Summary of Theoretical and Empirical Control Charts

---

THEORETICAL CONTROL CHARTS (STANDARDS GIVEN)

1. *Purpose:* Maintain surveillance of the mean, $\underset{\sim}{\mu}_0$, with Σ known, based on a sample of size n.

   Text Location: Chapter IV, Section 1.

   Test Statistic: $n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)$

   $UCL = \chi^2(p,\alpha)$, $LCL = 0$.

2. *Purpose:* Maintain surveillance of the variance-covariance matrix, Σ_0, based on a sample of size n.

   Text Location: Chapter VII, Section 2.

   a. *Test Statistic:* $-pn + pn\ \ell n n - n\ \ell n(|A|/|\Sigma_0|) + tr(\Sigma_0^{-1}A)$.

   $UCL = \chi^2\left[p(p+1)/2,\alpha\right]$, $LCL = 0$.

   b. *Test Statistic:* $|S|$

   $UCL = |\Sigma_0|(b_1+3b_2^{1/2})$, $CL = |\Sigma_0|b_1$, $LCL = |\Sigma_0|(b_1-3b_2^{1/2})$

   c. *Test Statistic:* $|S|^{1/2}$ (Multivariate Sigma Chart)

   $UCL = |\Sigma_0|(b_3+3b_4^{1/2})$, $CL = |\Sigma_0|^{1/2}b_3$, $LCL = |\Sigma_0|^{1/2}(b_3-3b_4^{1/2})$

   d. *Test Statistic:* $|S|$

   $UCL = |\Sigma_0|\left\{1+z_{\alpha/2}\sqrt{2p/(n-1)}\right\}$, $LCL = |\Sigma_0|\left\{1-z_{\alpha/2}\sqrt{2p/(n-1)}\right\}$

   e. *Test Statistic:* $|S|$

   $$UCL = \exp\{c_1 + p\ \ell n 2 + \sum_{h=1}^{p} \psi\left[(n-h)/2\right] + z_{\alpha/2}\sqrt{\sum_{h=1}^{p} \psi^{(1)}\left[(n-h)/2\right]}\}$$

   $$LCL = \exp\{c_1 + p\ \ell n 2 + \sum_{h=1}^{p} \psi\left[(n-h)/2\right] - z_{\alpha/2}\sqrt{\sum_{h=1}^{p} \psi^{(1)}\ (n-h)/2]}\}$$

Table 10. Continued

---

f. *Test Statistic:* $|S|$ (Valid only for $p = 2$)

$UCL = |\Sigma_0|\left(\chi^2(2n-4,\alpha/2)\right)^2/\left(4(n-1)^2\right)$

$LCL = |\Sigma_0|\left(\chi^2(2n-4,1-(\alpha/2))\right)^2/\left(4(n-1)^2\right)$

g. *Test Statistic:* $|S|^{1/2}$ (Valid only for $p = 2$)

$UCL = |\Sigma_0|^{1/2}\chi^2(2n-4,\alpha/2)/\left(2(n-1)\right)$

$LCL = |\Sigma_0|^{1/2}\chi^2\left(2n-4,1-(\alpha/2)\right)/\left(2(n-1)\right)$

EMPIRICAL CONTROL CHARTS (BASED ON PAST DATA)

3. *Purpose:* Analyze past data for lack of control of the mean based on k rational subgroups of n observations each.

   Text Location: Chapter V, Section 1.

   Test Statistic: $n(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})$

   $UCL = \left((knp-kp-np+p)/(kn-k-p+1)\right)F(p,kn-k-p+1,\alpha)$, $LCL = 0$

4. *Purpose:* Determine control of the process mean relative to μ_0 based on a sample of size m where S is computed for each sample of size m.

   Text Location: Chapter V, Section 2.

   Test Statistic: $m(\bar{\underset{\sim}{X}}_m-\mu_0)'S^{-1}(\bar{\underset{\sim}{X}}_m-\mu_0)$

   $UCL = \left(p(m-1)/(m-p)\right)F(p,m-p,\alpha)$, $LCL = 0$

5. *Purpose:* Determine control of the process mean relative to μ_0 based on a sample of size n where S_m is determined from a prior sample of size m.

   Text Location: Chapter V, Section 2.

   Test Statistic: $n(\bar{\underset{\sim}{X}}_n-\mu_0)'S_m^{-1}(\bar{\underset{\sim}{X}}_n-\mu_0)$

   $UCL = \left(p(m-1)/(m-p)\right)F(p,m-p,\alpha)$

Table 10. Continued

6. *Purpose:* Determine control of a process based on a sample of size n relative to $\bar{X}_m$ where S_m is also determined from a prior sample of size m.

Text Location: Chapter V, Section 2.

Test Statistic: $n(\bar{\underset{\sim}{X}}_n-\bar{\underset{\sim}{X}}_m)'S_m^{-1}(\bar{\underset{\sim}{X}}_n-\bar{\underset{\sim}{X}}_m)$

$UCL = [p(m+n)(m-1)/(m^2-mp)]F(p,m-p,\alpha)$, $LCL = 0$

7. *Purpose:* Analyze past data for lack of control of the dispersion based on k rational subgroups of n observations each.

Text Location: Chapter VII, Section 2.

a. *Test Statistic:* $-pn + pn\,\ell n n - n\ell n(|A^{(h)}|/(|S_p|b_5^{-1}) + tr(S_p^{-1}A^{(h)})$

$UCL = \chi^2[p(p+1)/2,\alpha]$, $LCL = 0$

b. *Test Statistic:* $|S^{(h)}|$

$UCL = (|S_p|b_5^{-1})(b_1+3b_2^{1/2})$, $CL = (|S_p|b_5^{-1})b_1$,

$LCL = (|S_p|b_5^{-1})(b_1-3b_2^{1/2})$

c. *Test Statistic:* $|S^{(h)}|^{1/2}$

$UCL = (|S_p|^{1/2}b_6^{-1})(b_3+3b_4^{1/2})$, $CL = (|S_p|^{1/2}b_6^{-1})b_3$

$LCL = (|S_p|^{1/2}b_6^{-1})(b_3-3b_4^{1/2})$

d. *Test Statistic:* $|S^{(h)}|$

$UCL = (|S_p|b_5^{-1})(1+z_{\alpha/2}\sqrt{2p/(n-1)})$, $LCL = (|S_p|b_5^{-1})(1-z_{\alpha/2}\sqrt{2p/(n-1)})$

e. *Test Statistic:* $|S^{(h)}|$

$$UCL = \exp\{c_2 + p\,\ell n 2 + \sum_{h=1}^{p}\psi[(n-h)/2] + z_{\alpha/2}\sqrt{\sum_{h=1}^{p}\psi^{(1)}[(n-h)/2]}\}$$

$$LCL = \exp\{c_2 + p\,\ell n 2 + \sum_{h=1}^{p}\psi[(n-h)/2] - z_{\alpha/2}\sqrt{\sum_{h=1}^{p}\psi^{(1)}[(n-h)/2]}\}$$

Table 10. Continued

f. *Test Statistic:* $|S^{(h)}|$ (Valid only for p = 2)

$$UCL = (|S_p|b_5^{-1})\left[\chi^2(2n-4,\alpha/2)\right]^2/\left[4(n-1)^2\right]$$

$$LCL = (|S_p|b_5^{-1})\left[\chi^2(2n-4,1-(\alpha/2))\right]^2/\left[4(n-1)\right]^2$$

g. *Test Statistic:* $|S^{(h)}|^{1/2}$ (Valid only for p = 2)

$$UCL = (|S_p|^{1/2}b_6^{-1})\chi^2(2n-4,\alpha/2)/\left[2(n-1)\right]$$

$$LCL = (|S_p|^{1/2}b_6^{-1})\chi^2\left[2n-4,1-(\alpha/2)\right]/\left[2(n-1)\right]$$

Although the concept of rational subgrouping for multivariate observations has been considered previously, the literature is in error. Based on the univariate procedure of replacing a population parameter by its unbiased estimate in the theoretical chart, a new statistic was proposed for the multivariate case and its distribution was developed and shown to be a special case of Hotelling's T^2 distribution. For each subgroup, simultaneous techniques were presented for determining those concepts responsible for the rejection. The power of the test statistic for each subgroup was also developed. An attempt was made to show how the multivariate chart reduces to the univariate chart. One avenue of future research is to obtain the joint distribution of the subgroup test statistics for all k subgroups. The utilization of correlations upon the component interval simultaneous techniques also deserves future consideration.

The charts numbered four, five, and six in Table 10 formed the basis for the remainder of Chapter V. All of these charts assumed that the population variance-covariance matrix was not known. The basic use of these charts was to determine the control of the process location relative to either some selected value $\underset{\sim}{\mu}_0$ or some value of the sample mean determined from prior data. For all of these charts, the power of the test and the use of simultaneous techniques was also presented.

Chapter VII dealt exclusively with control of the process dispersion. Chapter VII commenced with a review of the univariate theoretical and empirical control charts. For the case of rational subgroups, a new statistic was developed for testing the equality of variances and its distribution was obtained. One avenue of future research is to determine how these statistics for the k subgroups fare as competitors to other techniques for testing the equality of variances. Of course, this implies that the joint distribution of these k statistics also needs to be determined.

The charts under number two of Table 10 form the basis for the first half of Section 2 of Chapter VII. The chart, lettered a, merely utilized the test of significance viewpoint and borrowed upon this methodology. The other charts are based upon the sample generalized variance as a univariate measure of multivariate dispersion. Apparently, these charts have not previously appeared in the literature. The charts under number seven of Table 10 form the basis for the last half of Section 2 of Chapter VII. All of these charts are based upon the univariate technique of replacing the population parameters in the theoretical charts by their unbiased estimates.

Since all of the charts for both location and dispersion assumed that the parent population was normally distributed, an area of future research would be to determine the effect of nonnormality upon some of these procedures. For certain tests of significance, Ito (39) has done some work in this area. Another assumption for all of these charts was that the elements of the random sample were independent. Hoeffding and Robbins (32) have investigated a Central-Limit Theorem for dependent variables and this could be of some assistance for the charts on location. However, it seems that this would be of little assistance for the charts on dispersion. Thus, another area of future investigation would be the effect of nonrandomness among the elements of the sample.

Another major purpose of this research was to concisely present the statistical methodology needed for the various charts. A byproduct of this was the direct product proof of the independence of the sample mean vector and the sample variance-covariance matrix presented in Chapter III.

One additional byproduct of this research was the paradox and resolutions presented in Chapter VI. Since the paradox was treated exclusively for $p = 2$, an avenue of future research would be the extension of this paradox to $p \geq 3$.

APPENDIX

THE APPLICATION OF MULTIVARIATE CONTROL CHARTS

To illustrate the use of some of the multivariate control charts, data was generated from a bivariate normal distribution with the $\underset{\sim}{\mu}$ and Σ given in Equation (101). Thus, $\mu_1 = 15.0$, $\mu_2 = 5.0$, $\sigma_1 = 1.20$, $\sigma_2 = 0.80$, and $\rho = 0.70$. Actually, observations were generated from a univariate standard normal distribution, and the transformation $\underset{\sim}{X} = R\underset{\sim}{Z} + \underset{\sim}{\mu}$ was used, where $\Sigma = RR'$. The random variables X_1 and X_2 could represent the length and width, respectively, in inches of a steel plate.

$$\underset{\sim}{\mu} = \begin{bmatrix} 15.0 \\ 5.0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.440 & 0.672 \\ 0.672 & 0.640 \end{bmatrix} \tag{101}$$

1. Theoretical Control Charts

The first charts to be considered will be theoretical charts for location and dispersion. To obtain the data necessary for the use of these charts, 20 samples of 10 observations each were generated and these are given in Table 11. To illustrate the effect of a shift in the mean vector, μ_1 was set equal to 16, 17, and 16 inches for samples 10, 11, and 12, respectively. Note that even though the data for the 20 samples appear simultaneously in Table 11, the decision maker is only given one sample at a time, and a decision must be made after each sample. This is in contrast to the empirical charts to be discussed later.

Table 11. Twenty Samples with x_1 = Length and x_2 = Width of Steel Plates

(Sample Numbers Shown in Parentheses)

	x_1	x_2		x_1	x_2		x_1	x_2		x_1	x_2
(1)	14.76	5.06	(6)	13.12	4.26	(11)	17.50	4.97	(16)	15.91	5.52
	10.76	3.01		15.49	5.36		15.44	4.17		15.42	5.64
	15.35	5.05		14.44	4.37		17.16	5.23		19.91	7.02
	15.30	5.01		14.77	4.11		17.86	5.31		14.97	4.65
	14.24	4.58		16.47	6.29		18.73	6.71		15.01	4.69
	15.74	5.44		18.35	6.35		17.37	4.86		16.42	3.74
	16.23	5.32		16.80	6.04		17.39	4.89		15.61	5.84
	10.89	3.72		15.43	5.32		16.39	3.94		16.18	5.80
	15.32	5.05		15.43	5.89		17.96	5.43		14.52	4.24
	13.93	4.31		15.61	5.49		16.11	4.83		15.43	5.10
(2)	16.32	4.28	(7)	15.96	5.26	(12)	14.63	4.38	(17)	15.32	5.58
	15.28	5.01		16.49	5.76		15.53	4.66		14.19	3.37
	14.41	4.77		17.19	6.43		15.40	5.11		14.43	5.31
	18.52	7.51		14.47	4.44		16.65	5.72		14.84	5.14
	14.41	4.21		14.33	4.31		15.67	5.37		15.44	4.56
	14.07	4.46		14.36	4.34		14.87	4.62		13.80	4.73
	15.11	5.45		14.56	4.54		14.24	3.46		14.75	5.05
	15.12	4.89		14.95	4.90		16.19	5.31		15.86	4.40
	14.11	3.94		14.30	4.87		15.92	4.48		14.76	5.07
	18.07	7.11		15.03	5.56		15.82	4.39		13.83	4.77
(3)	13.81	4.24	(8)	15.94	5.85	(13)	13.69	4.47	(18)	15.47	4.61
	15.72	5.47		15.82	5.74		16.34	5.83		14.89	5.21
	16.61	4.03		14.68	5.24		19.17	6.79		14.48	4.83
	14.08	5.07		13.71	4.33		13.77	4.56		14.25	3.47
	15.32	5.10		14.12	4.15		15.75	5.28		15.40	5.70
	11.93	3.63		14.71	4.70		15.50	5.62		15.52	5.24
	14.73	4.55		14.27	4.29		14.23	4.43		14.62	4.40
	14.09	4.53		15.20	4.60		15.53	5.09		15.09	5.42
	13.63	3.53		15.11	5.09		18.21	7.62		16.94	6.59
	14.55	4.39		16.40	5.74		13.87	4.10		16.56	3.95
(4)	14.44	4.87	(9)	16.66	5.99	(14)	15.69	4.68	(19)	15.16	4.92
	13.31	3.80		15.89	5.27		16.50	4.87		15.13	4.90
	17.15	5.72		14.10	4.73		15.08	5.25		20.08	7.23
	13.97	4.43		14.89	4.90		16.23	5.77		15.60	5.35
	14.57	4.43		14.65	4.68		12.77	4.22		16.10	5.82
	15.34	5.73		15.79	5.19		15.47	5.06		14.37	7.06
	15.08	4.92		14.70	4.74		13.55	3.83		14.02	4.44
	12.60	4.30		14.99	5.01		14.21	3.31		15.05	5.41
	13.89	4.95		14.26	4.32		15.04	5.24		12.65	4.30
	15.37	4.63		14.90	4.36		14.85	5.06		15.23	5.02
(5)	14.61	3.92	(10)	15.51	4.57	(15)	19.63	6.72	(20)	14.38	4.79
	15.23	6.22		17.70	5.49		13.79	4.07		13.71	7.02
	14.83	4.70		18.87	6.02		15.33	5.52		14.41	4.83
	13.40	4.50		15.41	4.48		14.63	7.15		14.09	3.95
	15.75	5.58		15.72	4.21		14.12	4.39		15.14	4.38
	14.37	5.14		15.01	4.12		16.18	5.19		16.37	6.11
	13.77	3.72		15.68	4.75		11.21	3.37		12.97	4.63
	15.44	4.73		16.52	4.98		16.02	5.62		14.55	3.83
	14.89	5.35		16.34	4.81		13.91	4.11		13.90	3.80
	13.32	4.44		16.33	4.80		15.37	7.30		14.63	4.49

Theoretical Chart for the Mean, Σ Known

The first chart to be considered is the theoretical control chart for the mean when Σ is known. The vector $\underset{\sim}{\mu}_0$ has been specified by the decision maker and has the value given in Equation (101). Based on the presentation in Chapter IV, the value of $n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)'\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}_0)$ was computed for each sample, and these values are given in Table 12. With an α level of 0.05, the control limit is $\chi^2(2,.05) = 5.99$, and the control chart is illustrated in Figure 8. From Table 12 or Figure 8, it is seen that the test statistic for samples 10, 11, 12 and 16 plots out of control. For samples 10, 11, and 12, even though the smallest intentional increase in μ_1 was only 6.74%, the control chart easily detected this. When μ_1 was set equal to 17, which is an increase of 13.4%, the statistic plotted off the chart. Note that the statistic for sample 16 also plots out of control even though there was no shift in the population mean. This is not unreasonable in view of the .05 alpha level. When the sample statistic plots out of control, as it did for samples 10, 11, 12, and 16, the decision maker should obtain the Bonferroni intervals for μ_1 and μ_2 to assist in determining those components responsible for the rejection. The Bonferroni intervals, as given in Equation (16), are $\bar{x}_h \pm z_{\alpha/2p}(\sigma_h/\sqrt{n})$ for $h = 1,2$. Even though the test statistic for the other samples did not plot out of control, the Bonferroni intervals for these samples were also obtained and are presented in Table 12. For all of the samples for which the test statistic plotted in control, the Bonferroni intervals contain the population values of $\mu_1 = 15.0$ and $\mu_2 = 5.0$. For all of the samples for which the test statistic plotted

Table 12. Chi-Squared Values and Bonferroni Intervals for μ_1 and μ_2

Sample Number	Chi-Squared Value	$\bar{x}_1$	$\bar{x}_2$	Interval for μ_1	Interval for μ_2
1	3.891	14.252	4.655	[13.402,15.102]	[4.088,5.222]
2	2.279	15.541	5.163	[14.691,16.391]	[4.596,5.730]
3	4.678	14.448	4.453	[13.598,15.298]	[3.886,5.020]
4	1.290	14.572	4.776	[13.722,15.422]	[4.209,5.343]
5	1.165	14.593	4.830	[13.743,15.443]	[4.263,5.397]
6	2.582	15.590	5.349	[14.740,16.440]	[4.782,5.916]
7	0.231	15.165	5.039	[14.315,16.015]	[4.472,5.606]
8	0.020	14.996	4.972	[14.146,15.846]	[4.405,5.539]
9	0.493	15.085	4.919	[14.235,15.935]	[4.352,5.486]
10	30.929	16.309	4.823	[15.459,17.159]*	[4.256,5.390]
11	63.402	17.191	5.033	[16.341,18.041]*	[4.466,5.600]
12	8.721	15.492	4.750	[14.642,16.342]	[4.183,5.317]
13	2.843	15.607	5.379	[14.757,16.457]	[4.812,5.946]
14	1.844	14.940	4.728	[14.090,15.790]	[4.161,5.295]
15	3.512	15.010	5.343	[14.160,15.860]	[4.776,5.910]
16	7.496	15.937	5.224	[15.087,16.787]*	[4.657,5.791]
17	0.702	14.721	4.797	[13.871,15.571]	[4.230,5.364]
18	2.052	15.321	4.941	[14.471,16.171]	[4.374,5.508]
19	3.401	15.339	5.449	[14.489,16.189]	[4.882,6.016]
20	2.487	14.414	4.784	[13.564,15.264]	[4.217,5.351]

*The asterisked intervals do not contain the population values.

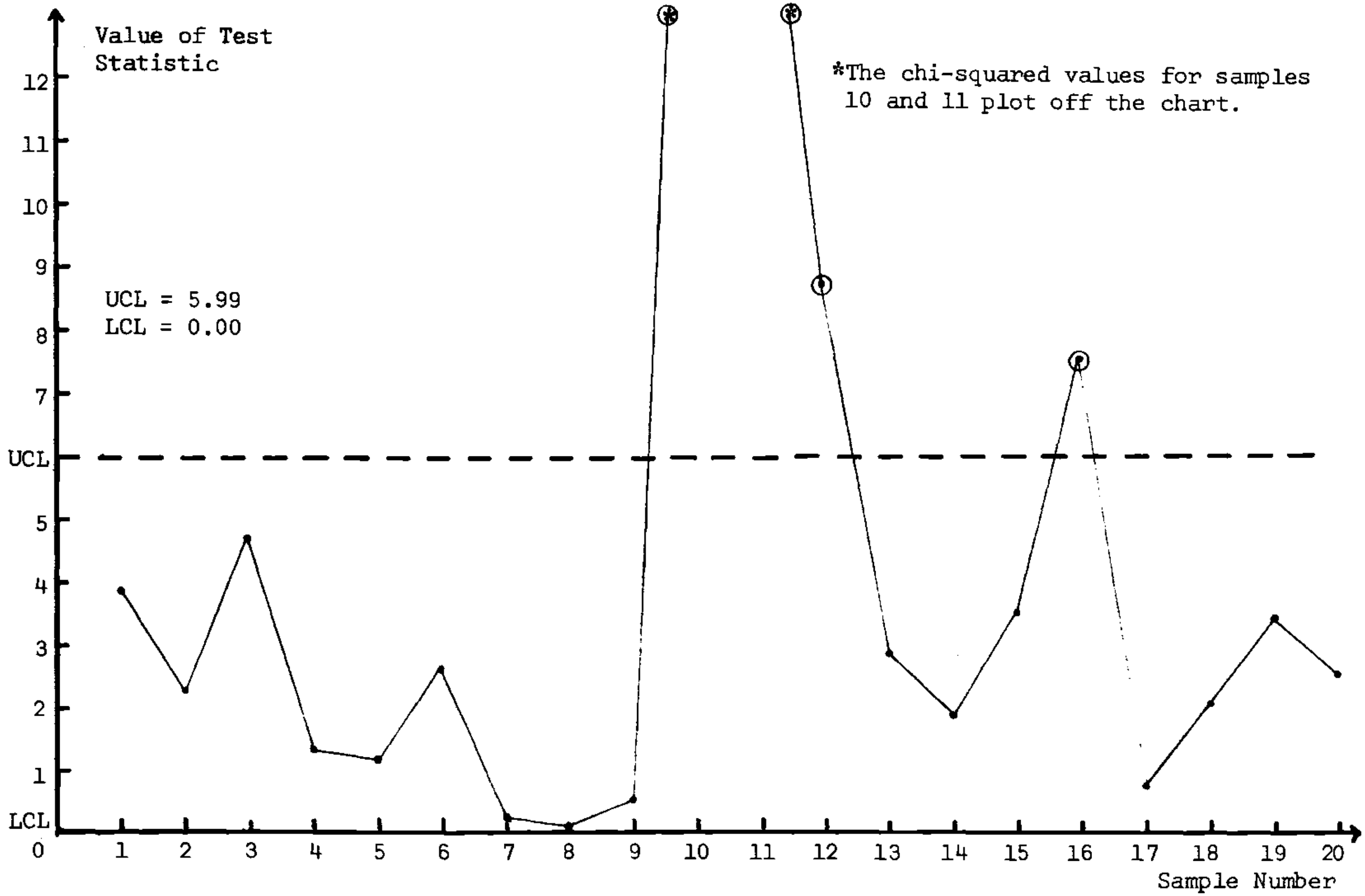

Figure 8. Theoretical Control Chart for the Mean with Σ Known

out of control, the Bonferroni intervals contain the population value of $\mu_2 = 5.0$. For these same samples, the intervals on μ_1 for samples numbered 10, 11, and 16 do not contain the population value of $\mu_1 = 15$. This should suggest to the decision maker that the first component, which is the length of the steel plate, is the basis for rejection, and an assignable cause should be sought. Note that even though the test statistic for sample number 12 plots out of control, the Bonferroni intervals would be of no assistance to the decision maker in determining the reason for rejection. This occurs since $\bar{\underset{\sim}{x}}$ for sample number 12 plots outside the elliptical control region, but within the rectangular region bounding the ellipse.

Theoretical Chart for Dispersion

For the data presented in Table 11, the control or lack of control of the process dispersion also needs to be determined. Thus, the next control chart to be considered is a theoretical chart for the dispersion where the population variance-covariance matrix (Σ_0) has been specified by the decision maker and has the value given in Equation (101). Although several theoretical dispersion charts were presented in Section 2 of Chapter VII for $|S|$, the one given by Equations (81) and (82) will be used since these limits were based on the exact distribution of $|S|$ for $p = 2$. Since $|\Sigma_0| = 0.47$, $\chi^2(16,.025) = 28.845$, and $\chi^2(16,.975) = 6.908$, it follows that UCL = 1.207 and LCL = 0.069. When S^2 is used to check process dispersion in univariate quality control, it is customary to use only an upper control limit with zero being the lower control limit. It would seem reasonable to follow suit for the multivariate extension. Thus, since S is a positive definite matrix with probability one, its determinant is greater than zero, and the lower control limit could be zero. In this case, UCL = $|\Sigma_0|\left(\chi^2(2n-4,.05)\right)^2/\left(4(n-1)^2\right) = 1.003$. The decision maker must determine which set of control limits to use. For each sample, the value of $|S|$ was computed and these are given in Table 13. If it is decided to adopt the control chart with UCL = 1.207 and LCL = 0.069, then the control chart based on this is illustrated in Figure 9. From Table 13 or Figure 9 it is seen that $|S|$ for samples 8, 9, and 10 plots below the lower control limit while $|S|$ for samples 15 and 19 plots above the upper control limit. It is important to realize that when the data were generated, there was no intentional increase in σ_1, σ_2, or ρ. Thus, the observed alpha level of .25 is five times greater than the specified level of .05. Even if the decision maker uses UCL = 1.003 and LCL = 0.000, the observed alpha level is .15 since now $|S|$ for samples 15, 16, and 19 plots above the UCL. It is suggested that the reason for the relatively large observed alpha levels stems from the failure of the data to conform to the requirements of normality or randomness, although there is the possibility that these requirements are met. In general, tests for dispersion are more sensitive to the underlying assumptions than tests for location.

To aid in the analysis of this control chart, the Bonferroni intervals were obtained for σ_1^2 and σ_2^2 for each of the 20 samples. For $h = 1,2$, let A_h be the event that $(\underline{\sigma}_h^2, \bar{\sigma}_h^2)$ covers σ_h^2, where $\underline{\sigma}_h^2 = (n-1)S_h^2/\chi^2(n-1,\gamma/2)$ and $\bar{\sigma}_h^2 = (n-1)S_h^2/\chi^2\left(n-1,1-(\gamma/2)\right)$. For $p = 2$ and a total

Table 13. Values of $|S|$ and Bonferroni Intervals for σ_1^2 and σ_2^2

Sample Number	Value of $\|S\|$	s_1^2	s_2^2	s_{12}	Interval for σ_1^2	Interval for σ_2^2
1	0.153	3.714	0.595	1.434	[1.589,15.064]*	[0.255,2.413]
2	0.804	2.567	1.479	1.730	[1.098,10.411]	[0.633,5.999]
3	0.496	1.647	0.403	0.409	[0.705,6.680]	[0.172,1.635]
4	0.232	1.606	0.365	0.595	[0.687,6.514]	[0.156,1.480]
5	0.249	0.705	0.579	0.399	[0.302,2.859]	[0.248,2.348]
6	0.374	2.000	0.706	1.019	[0.856,8.112]	[0.302,2.863]
7	0.086	1.056	0.496	0.662	[0.452,4.283]	[0.212,2.012]
8	0.064	0.746	0.422	0.501	[0.319,3.026]	[0.181,1.712]
9	0.031	0.631	0.238	0.345	[0.270,2.559]	[0.102,0.965]
10	0.029	1.379	0.329	0.652	[0.590,5.593]	[0.141,1.334]
11	0.135	0.939	0.580	0.640	[0.402,3.808]	[0.248,2.352]
12	0.088	0.539	0.422	0.374	[0.231,2.186]	[0.181,1.712]
13	0.528	3.558	1.267	1.995	[1.522,14.431]*	[0.542,5.139]
14	0.411	1.354	0.551	0.579	[0.579,5.492]	[0.236,2.235]
15	4.959	4.697	1.898	1.989	[2.010,19.050]*	[0.812,7.698]*
16	1.104	2.279	0.887	0.958	[0.975,9.243]	[0.380,3.598]
17	0.172	0.468	0.377	0.066	[0.200,1.898]	[0.161,1.529]
18	0.469	0.753	0.804	0.370	[0.322,3.054]	[0.344,3.261]
19	1.963	3.684	1.021	1.341	[1.576,14.942]*	[0.437,4.141]
20	0.843	0.818	1.065	0.169	[0.350,3.318]	[0.456,4.320]

*The asterisked intervals do not contain the population values.

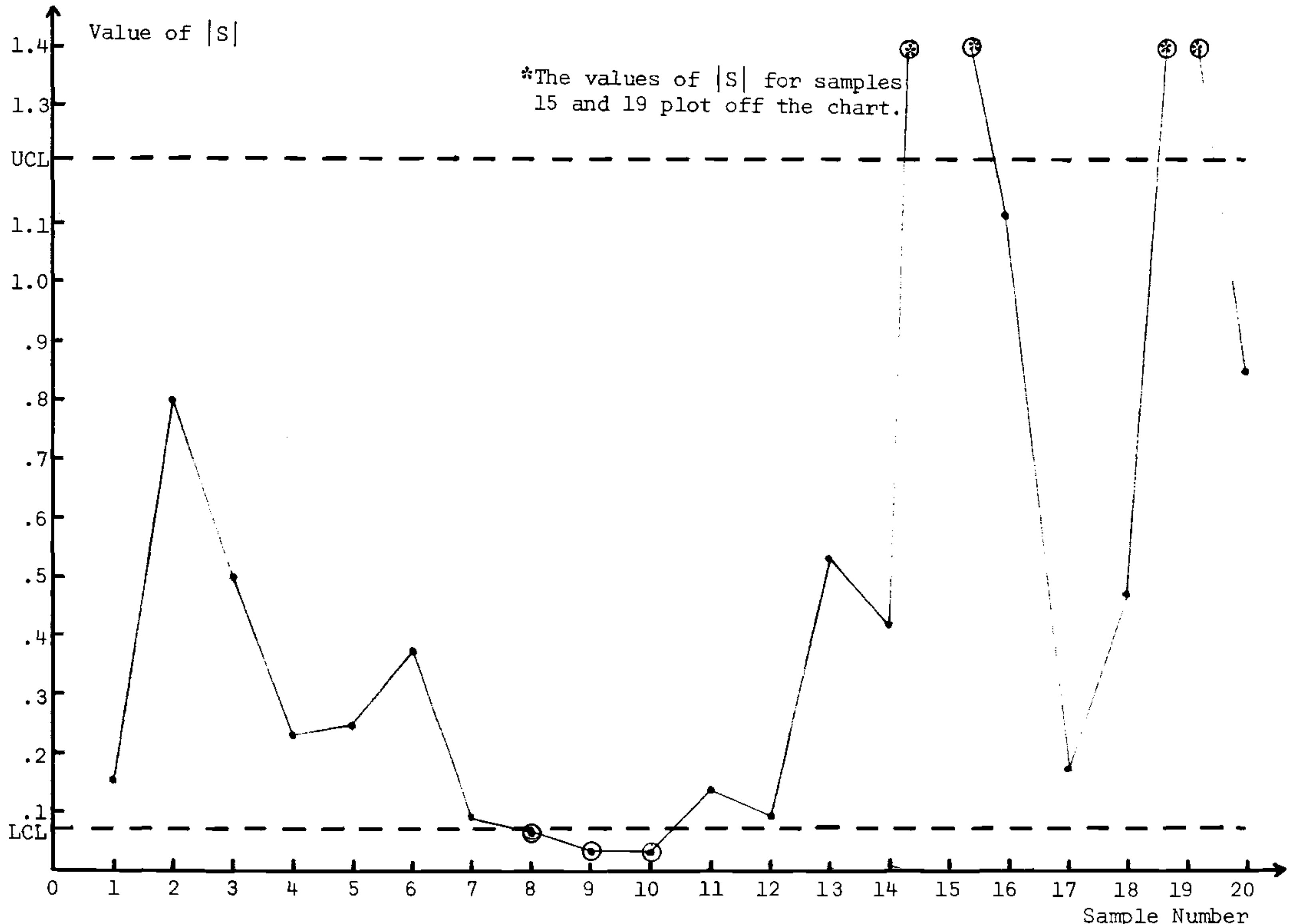

Figure 9. Theoretical Control Chart for the Dispersion

alpha of .05, $\gamma/2 = .0125$ and $1 - (\gamma/2) = .9875$. Thus, $\chi^2(9,.0125) = 21.034$ and $\chi^2(9,.9875) = 2.219$. The Bonferroni intervals are given in Table 13. For samples 1, 13, 15, and 19, the intervals for σ_1^2 do not contain the population value of $\sigma_1^2 = 1.440$. In addition, the interval on σ_2^2 for sample 15 does not contain the population value of $\sigma_2^2 = 0.640$. Thus, when $|S|$ for sample number 15 plots out of control, the Bonferroni intervals suggest that the variances of both the length and the width have gone astray; when $|S|$ for sample number 19 plots out of control, the Bonferroni intervals suggest that this is caused only by the variance of the second component. Also, even though $|S|$ for samples 1 and 13 plots in control, the Bonferroni intervals indicate that the variance of the first component has gone astray. The Bonferroni intervals are of no assistance for those samples which plotted below the lower control limit since the Bonferroni intervals contain the population values of both σ_1^2 and σ_2^2 for these samples.

A further analysis of the control chart for $|S|$ is presented in Figure 10, which contains bivariate data plots for samples numbered 1, 8, 9, 10, 13, 15, and 19. This illustrates the fact that $|S|$ is a linear measure of dispersion. That is, the greater the departure of the data from lying on a straight line, the greater the value of $|S|$. These data plots confirm that the $|S|$ for samples 15 and 19 should indeed plot above the upper control limit. Also, the $|S|$ for samples 8, 9, and 10 plots below the lower control limit in view of the lack of both linear dispersion and overall dispersion. Finally, the $|S|$ for samples numbered 1 and 13 plots in control because of the small linear dispersion,

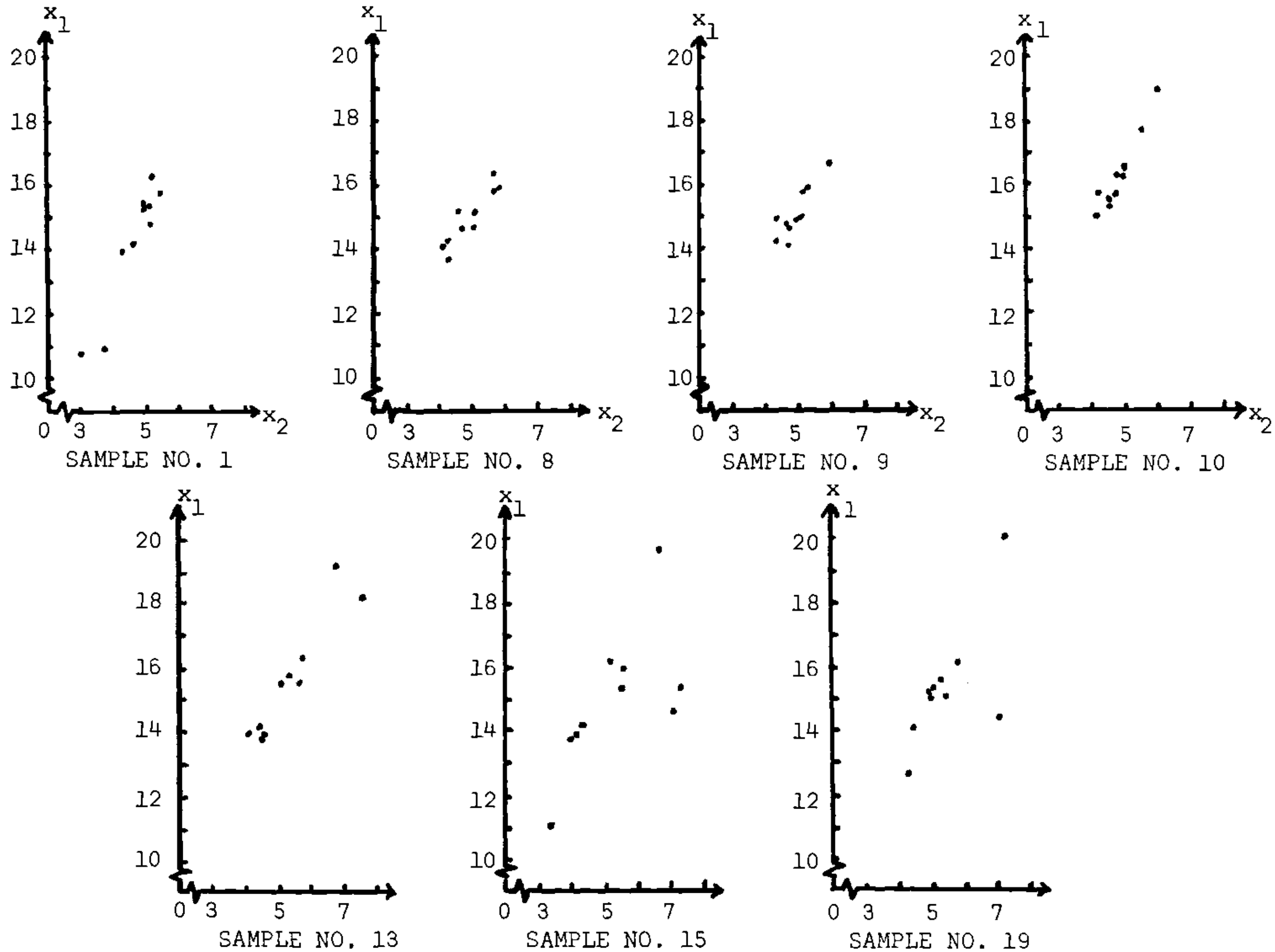

Figure 10. Bivariate Plots of X_1 and X_2

yet the relatively large dispersion of the first component precludes the possibility of the Bonferroni intervals bracketing the population value of σ_1^2. This suggests that the control chart should always be complemented by the Bonferroni intervals.

As a final point of interest, values were obtained for the control limits which were based on various approximations to the distribution of $|S|$. From Equations (73) and (74), with $\alpha = .05$, it follows that UCL = 1.084 and LCL = -0.144. Since the determinant of S is always nonnegative, the lower control limit is replaced by 0. From Equations (77) and (78), it follows that UCL = 1.368 and LCL = 0.078. These limits compare quite favorably with the exact limits, even though $p = 2$ and $n = 10$.

2. Empirical Control Charts

The next set of charts to be considered will be empirical charts for location and dispersion. To obtain the data necessary for the use of these charts, 25 samples of 10 observations each were generated and these are given in Table 14. Here, there was no intentional alteration of any of the population parameters. Thus, except for samples 10, 11, and 12 and the last five samples, the data given here is identical with that given in Table 11. In the construction of empirical control charts, the decision maker has all 25 sets of data prior to the determination of the control limits. Also, the population parameters specified by Equation (101) are unknown and must be estimated from the 25 data sets. These estimates are given in Equation (102).

Table 14. Twenty-Five Samples with x_1 = Length and x_2 = Width of Steel Plates

(Samples Numbers Shown in Parentheses)

	x_1	x_2		x_1	x_2		x_1	x_2		x_1	x_2		x_1	x_2
(1)	14.76	5.06	(6)	13.12	4.26	(11)	15.50	4.97	(16)	15.91	5.52	(21)	13.14	4.80
	10.76	3.01		15.49	5.36		13.44	4.17		15.42	5.64		15.42	4.67
	15.35	5.05		14.44	4.37		15.16	5.23		19.91	7.02		14.27	4.74
	15.30	5.01		14.77	4.11		15.86	5.31		14.97	4.65		14.51	3.81
	14.24	4.58		16.47	6.29		16.73	6.71		15.01	4.69		16.11	5.33
	15.74	5.44		18.35	6.35		15.37	4.86		16.42	3.74		15.49	5.32
	16.23	5.32		16.80	6.04		15.39	4.89		15.61	5.84		18.65	6.59
	10.89	3.72		15.43	5.32		14.39	3.94		16.18	5.80		15.98	5.79
	15.32	5.05		15.43	5.89		15.96	5.43		14.52	4.24		14.69	4.58
	13.93	4.31		15.61	5.49		14.11	4.83		15.43	5.10		15.98	6.36
(2)	16.32	4.28	(7)	15.96	5.26	(12)	13.63	4.38	(17)	15.32	5.58	(22)	15.04	4.91
	15.28	5.01		16.49	5.76		14.53	4.66		14.19	3.37		15.47	5.89
	14.41	4.77		17.19	6.43		14.40	5.11		14.43	5.31		17.28	5.89
	18.52	7.51		14.47	4.44		15.65	5.72		14.84	5.14		16.86	6.07
	14.41	4.21		14.33	4.31		14.67	5.37		15.44	4.56		14.22	4.73
	14.07	4.46		14.36	4.34		13.87	4.62		13.80	4.73		14.16	4.10
	15.11	5.45		14.56	4.54		13.24	3.46		14.75	5.05		15.47	4.77
	15.12	4.89		14.95	4.90		15.19	5.31		15.86	4.40		14.56	5.05
	14.11	3.94		14.30	4.87		14.92	4.48		14.76	5.07		13.82	4.36
	18.07	7.11		15.03	5.56		14.82	4.39		13.83	4.77		14.46	3.82
(3)	13.81	4.24	(8)	15.94	5.85	(13)	13.69	4.47	(18)	15.47	4.61	(23)	15.27	5.74
	15.72	5.47		15.82	5.74		16.34	5.83		14.89	5.21		16.26	6.67
	16.61	4.03		14.68	5.24		19.17	6.79		14.48	4.83		15.02	5.51
	14.08	5.07		13.71	4.33		13.77	4.56		14.25	3.47		15.16	5.07
	15.32	5.10		14.12	4.15		15.75	5.28		15.40	5.70		14.27	4.81
	11.93	3.63		14.71	4.70		15.50	5.62		15.52	5.24		14.76	5.28
	14.73	4.55		14.27	4.29		14.23	4.43		14.62	4.40		14.23	4.20
	14.09	4.53		15.20	4.60		15.53	5.09		15.09	5.42		15.07	5.00
	13.63	3.53		15.11	5.09		18.21	7.62		16.94	6.59		13.68	4.27
	14.55	4.39		16.40	5.74		13.87	4.10		16.56	3.95		16.07	5.38
(4)	14.44	4.87	(9)	16.66	5.99	(14)	15.69	4.68	(19)	15.16	4.92	(24)	17.44	6.10
	13.31	3.80		15.89	5.27		16.50	4.87		15.13	4.90		16.58	5.87
	17.15	5.72		14.10	4.73		15.08	5.25		20.08	7.28		14.70	4.67
	13.97	4.43		14.89	4.90		16.23	5.77		15.60	5.35		15.39	5.32
	14.57	4.43		14.65	4.68		12.77	4.22		16.10	5.82		15.06	5.01
	15.34	5.73		15.79	5.19		15.47	5.06		14.37	7.06		13.70	4.31
	15.08	4.92		14.70	4.74		13.55	3.83		14.02	4.44		16.12	5.45
	12.60	4.30		14.99	5.01		14.21	3.31		15.05	5.41		16.31	5.64
	13.89	4.95		14.26	4.32		15.04	5.24		12.65	4.30		14.28	4.87
	15.37	4.63		14.90	4.36		14.85	5.06		15.23	5.02		14.83	4.24
(5)	14.61	3.92	(10)	14.51	4.57	(15)	19.63	6.72	(20)	14.38	4.79	(25)	15.55	4.92
	15.23	6.22		16.70	5.49		13.79	4.07		13.71	7.02		15.24	5.21
	14.83	4.70		17.87	6.02		15.33	5.52		14.41	4.83		13.91	4.53
	13.40	4.50		14.41	4.48		14.63	7.15		14.09	3.95		13.96	4.58
	15.75	5.58		14.72	4.21		14.12	4.39		15.14	4.38		15.38	5.35
	14.67	5.14		14.01	4.12		16.18	5.19		16.37	6.11		15.78	4.59
	13.77	3.72		14.68	4.75		11.21	3.37		12.97	4.63		15.15	4.57
	15.44	4.73		15.52	4.98		16.02	5.62		14.55	3.83		14.70	4.72
	14.89	5.35		15.34	4.81		13.81	4.11		13.90	3.80		15.62	5.59
	13.32	4.44		15.33	4.80		15.37	7.30		14.63	4.49		16.72	6.06

$$\underset{\sim}{\bar{\bar{X}}} = \begin{bmatrix} 15.068 \\ 4.997 \end{bmatrix}, \quad S_p = \begin{bmatrix} 1.653 & 0.777 \\ 0.777 & 0.681 \end{bmatrix} \tag{102}$$

Empirical Chart for the Mean, Σ Unknown

The first empirical chart to be considered is a chart for the mean when Σ is unknown. Based on the presentation in Chapter V, the value of $n(\underset{\sim}{\bar{X}}^{(h)}-\underset{\sim}{\bar{\bar{X}}})'S_p^{-1}(\underset{\sim}{\bar{X}}^{(h)}-\underset{\sim}{\bar{\bar{X}}})$ was computed for $h = 1,2,\ldots,25$, and these values are given in Table 15. With $\alpha = .05$, $k = 25$, and $n = 10$, the control limit is

$$\left[(k^2np-k^2p-knp+kp)/(k^2n-k^2-kp+k)\right]F(p,kn-k-p+1,.05) = 5.863,$$

and the control chart is illustrated in Figure 11. From Table 15 or Figure 11, it is seen that no values of the test statistic plot out of control. Apparently, the process is in control.

Empirical Chart for Dispersion

For the data presented in Table 14, the control or lack of control of the process dispersion also needs to be determined. Thus, the next control chart to be considered in an empirical chart for dispersion, where neither $\underset{\sim}{\mu}$ nor Σ are known. Although several empirical dispersion charts were presented in Section 2 of Chapter VII for $|S^{(h)}|$, the one given by Equations (97) and (98) will be used. Hence, for $\alpha = .05$,

$$UCL = (|S_p|b_5^{-1})\left(\chi^2(2n-4,.025)\right)^2/\left(4(n-1)^2\right),$$

Table 15. Values of $n(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})'S_p^{-1}(\bar{\underset{\sim}{X}}^{(h)}-\bar{\bar{\underset{\sim}{X}}})$

Sample No.	Value of Statistic	$\bar{x}_1$	$\bar{x}_2$
1	4.090	14.252	4.655
2	1.455	15.541	5.163
3	4.346	14.448	4.453
4	1.494	14.572	4.776
5	1.470	14.593	4.830
6	2.006	15.590	5.349
7	0.057	15.165	5.039
8	0.034	14.996	4.972
9	0.235	15.085	4.919
10	2.969	15.309	4.823
11	0.109	15.191	5.033
12	2.032	14.492	4.750
13	2.288	15.607	5.379
14	1.477	14.940	4.728
15	4.445	15.010	5.343
16	5.616	15.937	5.224
17	0.772	14.721	4.797
18	1.357	15.321	4.941
19	3.802	15.339	5.449
20	2.872	14.414	4.784
21	0.812	15.425	5.201
22	0.170	15.134	4.960
23	1.831	14.979	5.192
24	0.860	15.441	5.149
25	0.172	15.200	5.013

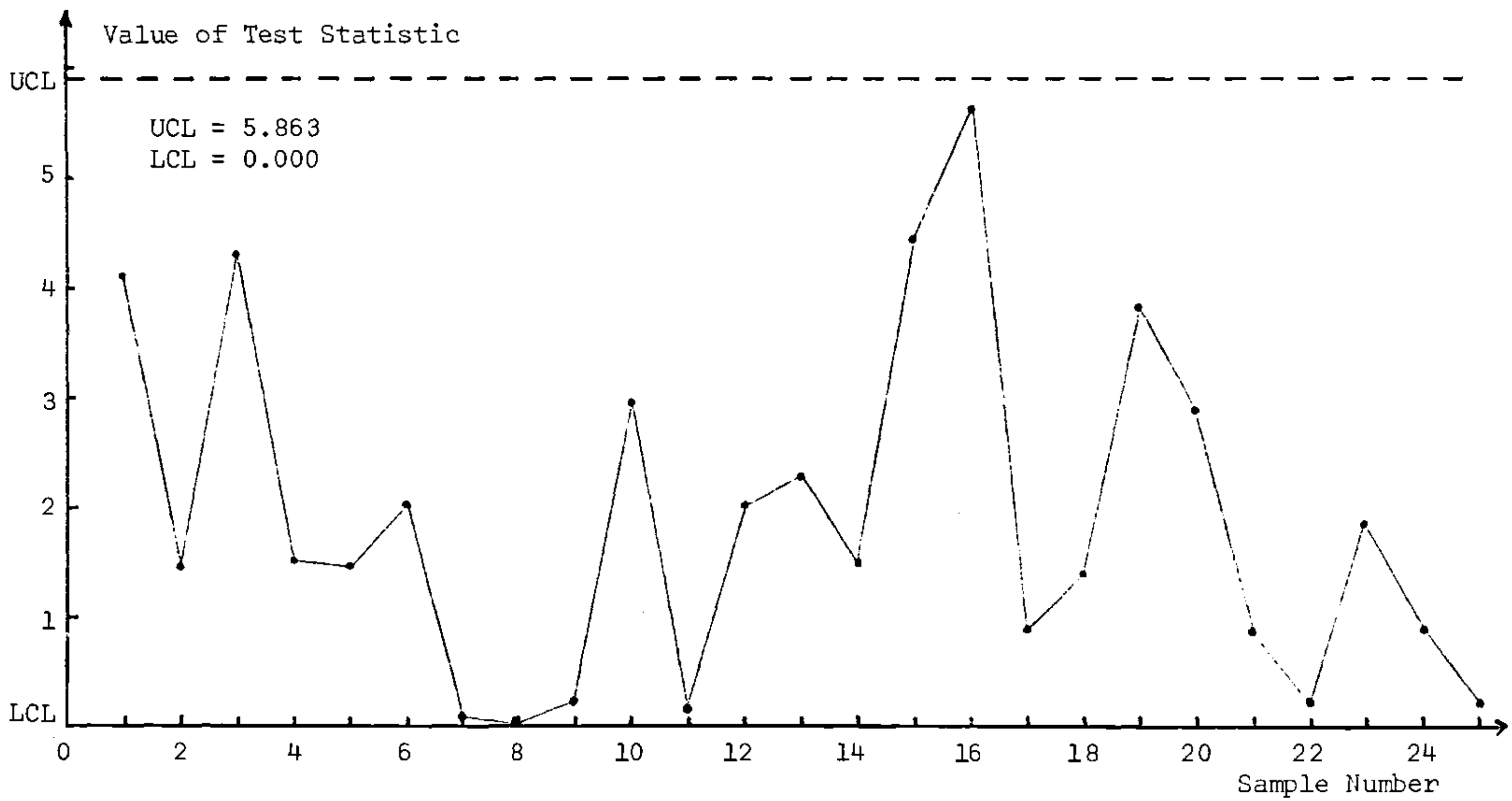

Figure 11. Empirical Control Chart for the Mean

$$LCL = (|S_p|b_5^{-1})\left(\chi^2(2n-4,.975)\right)^2/\left(4(n-1)^2\right),$$

and

$$b_5 = k^{-p}(n-1)^{-p} \prod_{h=1}^{p} (kn-k+1-h).$$

The constant b_5 is a correction factor for bias. Since $k = 25$, $n = 10$, and $p = 2$, $b_5 = .996$, which suggests that there is a minimal correction for bias. From Equation (102), it follows that $|S_p| = 0.522$. Thus, for the data presented in Table 14, the unbiased estimate of $|\Sigma_0|$ is $|S_p|b_5^{-1} = 0.524$, and UCL = 1.346 while LCL = 0.077. The control chart is illustrated in Figure 12, and the values of $|S^{(h)}|$ are given in Table 16. From Figure 12 or Table 16, it is seen that $|S^{(h)}|$ for samples 8, 9, and 10 plots below the lower control limit while, for samples 15 and 19, $|S^{(h)}|$ plots above the upper control limit. The same analysis applies here as for the theoretical dispersion chart.

Table 16. Values of $|S^{(h)}|$

Sample Number	Value of $\|S^{(h)}\|$	s_1^2	s_2^2	s_{12}
1	0.153	3.714	0.595	1.434
2	0.804	2.567	1.479	1.730
3	0.496	1.647	0.403	0.409
4	0.232	1.606	0.365	0.595
5	0.249	0.705	0.579	0.399
6	0.374	2.000	0.706	1.019
7	0.086	1.056	0.496	0.662
8	0.064	0.746	0.422	0.501
9	0.031	0.631	0.238	0.345
10	0.029	1.379	0.329	0.652
11	0.135	0.939	0.580	0.640
12	0.088	0.539	0.422	0.374
13	0.528	3.558	1.267	1.995
14	0.411	1.354	0.551	0.579
15	4.959	4.697	1.898	1.989
16	1.104	2.279	0.887	0.958
17	0.172	0.468	0.377	0.066
18	0.469	0.753	0.804	0.370
19	1.963	3.684	1.021	1.341
20	0.843	0.818	1.065	0.169
21	0.651	2.160	0.738	0.971
22	0.259	1.347	0.607	0.747
23	0.088	0.639	0.517	0.492
24	0.079	1.329	0.405	0.678
25	0.088	0.713	0.275	0.328

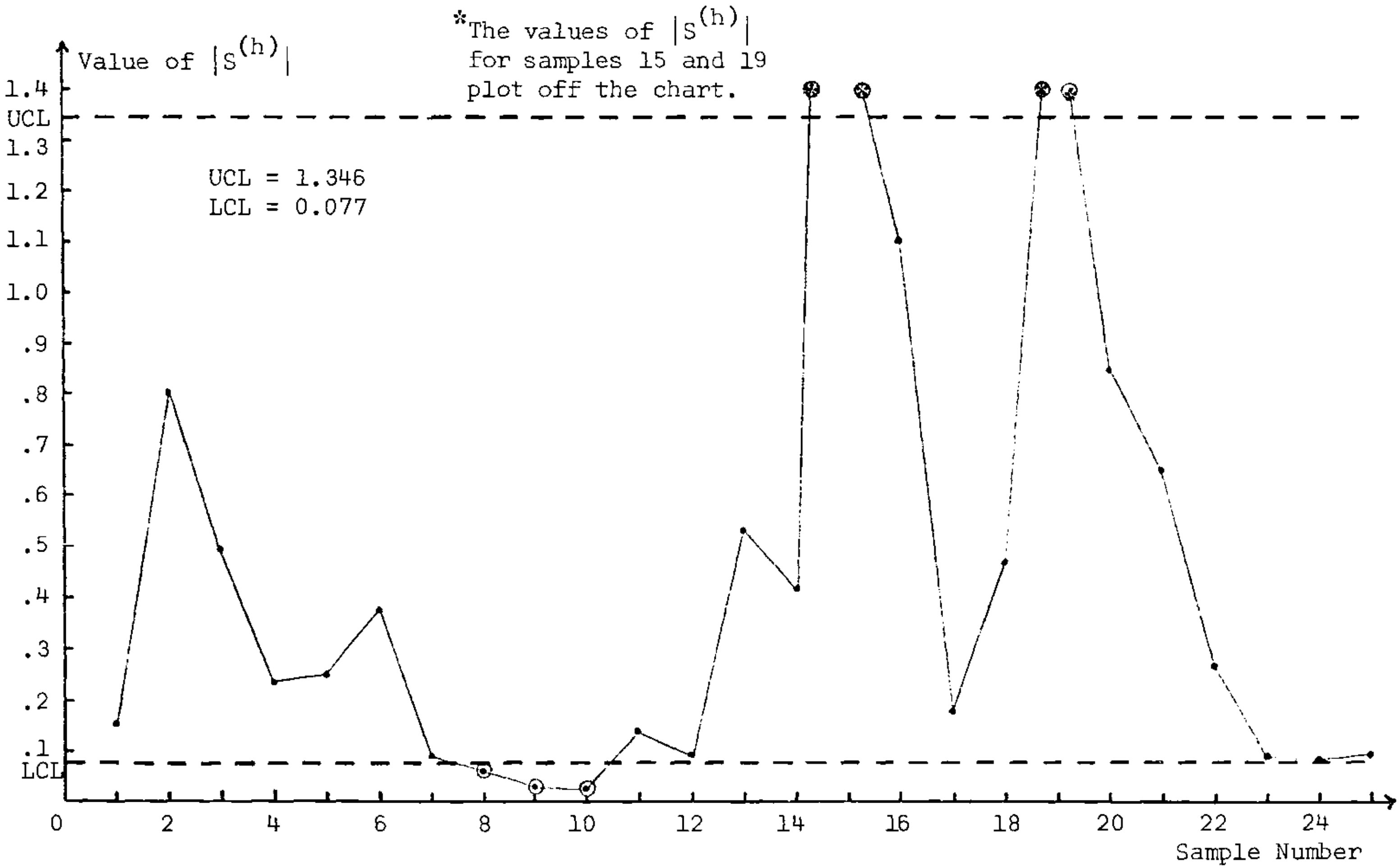

Figure 12. Empirical Control Chart for Dispersion

BIBLIOGRAPHY

1. Abramowitz, M. and Stegun, I. A. (Ed.), *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series No. 55, Washington, D.C.: U. S. Government Printing Office, 1964.

2. Aitken, A. C., "On the Wishart distribution in statistics," *Biometrika*, 36, 59-62, 1949.

3. Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc., New York, 1958.

4. Bartlett, M. S., "Properties of sufficiency and statistical tests," *Proc. Roy. Soc.*, Ser. A, 160, 268-282, 1937.

5. Bartlett, M. S., "Further aspects of the theory of multiple regression," *Proc. Camb. Phil. Soc.*, 34, 33-40, 1938.

6. Bartlett, M. S. and Kendall, D. G., "The statistical analysis of variance heterogeneity and the logarithmic transformation," *Journal of the Royal Statistical Society*, Ser. B, 8, 128-138, 1946.

7. Bowker, A. H., "A representation of Hotelling's T^2 and Anderson's classification statistic W in terms of simple statistics," in I. Olkin et al. (eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pp. 142-149, Stanford University Press, Stanford, Calif., 1960.

8. Bowker, A. H. and Lieberman, G. J., *Engineering Statistics*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1959.

9. Box, G. E. P., "A general distribution theory for a class of likelihood criteria," *Biometrika*, 36, 317-346, 1949.

10. Box, G. E. P., "Non-normality and tests on variances," *Biometrika*, 40, 318-335, 1953.

11. Boyd, W. C., "A nomogram for chi-square," *Journal of the American Statistical Association*, 60, 344-346, 1965.

12. Constantine, A. G., "The distribution of Hotelling's generalised T_0^2," *Ann. Math. Statist.*, 37, 215-225, 1966.

13. Craig, A. T., "Note on the Independence of Certain Quadratic Forms," *Ann. Math. Statist.*, 14, 195-197, 1943.

14. Craig, C. C., "Some remarks concerning the lot plot plan," *Industrial Quality Control*, September, 41-48, 1953.

15. Cramér, H., *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.

16. David, H. A., *Order Statistics*, John Wiley & Sons, Inc., New York, 1970.

17. Davis, H. T., *Tables of the Higher Mathematical Functions*, 2 Vols., Principia Press, Bloomington, Indiana, 1933, 1935.

18. Duncan, A. J., *Quality Control and Industrial Statistics*, Richard D. Irwin, Inc., Homewood, Illinois, 1965.

19. Dunn, O. J., "Estimation of the means of dependent variables," *Ann. Math. Statist.*, 29, 1095-1111, 1958.

20. Dunn, O. J., "Confidence intervals for the means of dependent, normally distributed variables," *Journal of the American Statistical Association*, 54, 613-621, 1959.

21. Esary, J. D., Proschan, F. and Walkup, D. W., "Association of random variables, with applications," *Ann. Math. Statist.*, 38, 1466-1474, 1967.

22. Fix, Evelyn, "Tables of non-central χ^2," *University of California Publications in Statistics*, Vol. 1, No. 2, 15-19, 1949.

23. Ghare, P. M. and Torgersen, P. E., "The Multicharacteristic Control Chart," *The Journal of Ind. Eng.*, Vol. XIX, No. 6, 269-272, 1968.

24. Gnanadesikan, M. and Gupta, S. S., "A selection procedure for multivariate normal distributions in terms of the generalized variances," *Technometrics*, 12, 103-117, 1970.

25. Graybill, F. A., *An Introduction to Linear Statistical Models*, Volume I, McGraw-Hill Book Company, Inc., New York, 1961.

26. Graybill, F. A., *Introduction to Matrices with Applications in Statistics*, Wadsworth Pub. Co., Belmont, Calif., 1969.

27. Guttman, I. and Wilks, S. S., *Introductory Engineering Statistics*, John Wiley & Sons, Inc., New York, 1965.

28. Harter, H. L. and D. B. Owen (Ed.), *Selected Tables in Mathematical Statistics*, Volume I, Markham Publishing Co., Chicago, 1970.

29. Haynam, G. E., Govindarajulu, Z. and Leone, F. C., "Tables of the cumulative non-central chi-square distribution," in H. Harter

et al. (eds.), *Selected Tables in Mathematical Statistics*, pp. 1-78, Markham Publishing Co., Chicago, 1970.

30. Hicks, C. R., "Some applications of Hotelling's T," *Industrial Quality Control*, June 23-26, 1955.

31. Hill, G. W., "Progress results on asymptotic approximations for Student's t and chi-squared, personal communication," in E. A. Cornish, *Fisher Memorial Lecture*, 37th I.S.I. Session, London, 1969.

32. Hoeffding, W. and Robbins, H., "The central limit theorem for dependent random variables," *Duke Math. Journal*, 15, 773-780, 1948.

33. Hoel, P. G., "A significance test for component analysis," *Ann. Math. Statist.*, 8, 149-158, 1937.

34. Hogg, R. V. and A. T. Craig, *Introduction to Mathematical Statistics*, The Macmillan Co., New York, 1965.

35. Hotelling, H., "The generalization of Student's ratio," *Ann. Math. Statist.*, 2, 360-378, 1931.

36. Hotelling, H., "Multivariate quality control, illustrated by the air testing of sample bombsights," *Techniques of Statistical Analysis*, McGraw-Hill Book Co., Inc., New York, 111-184, 1947.

37. Hotelling, H., "A generalized T test and measure of multivariate dispersion," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Univ. of Calif. Press, Berkeley, 23-41, 1951.

38. Hsu, P. L., "Notes on Hotelling's generalized T," *Ann. Math. Statist.*, 9, 231-243, 1938.

39. Ito, K., "On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures," in P. R. Krishnaiah (ed.), *Multivariate Analysis*, Vol. 2, p. 87-120, Academic Press, New York, 1968.

40. Jackson, J. E., "Quality control methods for two related variables," *Industrial Quality Control*, January, 4-8, 1956.

41. Jackson, J. E., "Quality control methods for several related variables," *Technometrics*, 1, 359-377, 1959.

42. Jackson, J. E. and Morris, R. H., "An application of multivariate quality control to photographic processing," *Journal of the American Statistical Association*, 52, 186-199, 1957.

43. James-Levy, G. E., "A nomogram for the integral law of Student's distribution," *Teoriya Veroyatnostei i ee Primeneniya,* 1, 271-274 (in Russian) (pp. 246-248 in English Translation).

44. Johnson, N. L. and Kotz, S., *Discrete Distributions,* Houghton Mifflin Company, Boston, 1969.

45. Johnson, N. L. and Kotz, S., *Continuous Univariate Distributions-1,* Houghton Mifflin Company, Boston, 1970.

46. Johnson, N. L. and Leone, F. C., *Statistics and Experimental Design in Engineering and the Physical Sciences,* Volume I, John Wiley & Sons, Inc., New York, 1964.

47. Kendall, M. G. and Stuart, A., *The Advanced Theory of Statistics,* Vol. 2, Hafner Publishing Co., New York, 1967.

48. Korin, B. P., "On the distribution of a statistic used for testing a covariance matrix," *Biometrika,* 55, 171-178, 1968.

49. Kshirsagar, A. M., *Multivariate Analysis,* Marcel Dekker, Inc., New York, 1972.

50. Lachenbruch, P. A., "The noncentral F distribution--extensions of Tang's tables," *Ann. Math. Statist.,* 37, 774 (Abstract), 1966.

51. Lancaster, H. O., *The Chi Squared Distribution,* John Wiley & Sons, Inc., New York, 1969.

52. Lindgren, B. W., *Statistical Theory,* The Macmillan Co., New York, 1969.

53. Miller, R. G., *Simultaneous Statistical Inference,* The McGraw-Hill Book Co., New York, 1966.

54. Montgomery, D. C. and Wadsworth, H. M., "Some techniques for multivariate quality control applications," *Transactions of the A.S.Q.C.,* 1972.

55. Morrison, D. F., *Multivariate Statistical Methods,* The McGraw-Hill Book Co., New York, 1967.

56. Pearson, E. S. and Hartley, H. O., "Charts of the power function for analysis of variance tests, derived from the noncentral F-distribution," *Biometrika,* 38, 112-130, 1951.

57. Rao, C. R., "An asymptotic expansion of the distribution of Wilk's criterion," *Bull. Int. Stat. Inst.,* 33(2), 177-180, 1951.

58. Rao, C. R., *Linear Statistical Inference and Its Applications,* John Wiley & Sons, Inc., 1965.

59. Resnikoff, G. J. and Lieberman, G. J., *Tables of the non-central t-distribution*, Stanford University Press, Stanford, Calif., 1957.

60. Roy, S. N., *Some Aspects of Multivariate Analysis*, John Wiley & Sons, Inc., New York, 1957.

61. Schatzoff, M., "Exact distributions of Wilk's likelihood ratio criterion," *Biometrika*, 53, 347-358, 1966.

62. Scheffé, H., "A method for judging all contrasts in the analysis of variance," *Biometrika*, 40, 87-104, 1953.

63. Scheffé, H., *The Analysis of Variance*, John Wiley & Sons, Inc., New York, 1959.

64. Silvey, S. D., *Statistical Inference*, Penguin Books, Baltimore, Maryland, 1970.

65. Siotani, M., "On the range in the multivariate case," *Proc. Inst. Statist. Math.*, 6, 155-165 (in Japanese), 1959.

66. Siotani, M., "The extreme value of the generalized distances of the individual points in the multivariate normal sample," *Ann. Inst. Statist. Math.*, Tokyo, 10, 183-203, 1959.

67. Tiku, M. L., "Tables of the power of the F-test," *Journal of the American Statistical Association*, 62, 525-539, 1967.

68. Tucker, *An Introduction to Probability and Mathematical Statistics*, Academic Press, New York, 1962.

69. Vaart, H. R., van der, "A note on the derivation of the non-central chi-square density function," *Statistica Neerlandica*, 21, 99-100, 1967.

70. Wilks, S. S., "Certain generalizations in the analysis of variance," *Biometrika*, 24, 471-494, 1932.

71. Wilks, S. S., *Mathematical Statistics*, John Wiley & Sons, Inc., New York, 1962.

72. Wishart, John, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, 20A, 32-52, 1928.

73. Wishart, John, "Proofs of the distribution law of the second order moment statistics," *Biometrika*, 35, 55-57, 1948.